Luglio 1992
Numero 287
Anno XXV
Volume XLVIII

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

Modelli matematici della diffusione dell'AIDS

Recenti simulazioni al calcolatore consentono di dare impreviste e originali risposte a due fra le più dibattute questioni sollevate dalla pandemia di AIDS: come si è propagata e come può essere rallentata

di Roy M. Anderson e Robert M. May

Nonostante dieci anni di ricerche intensive e di sforzi continui per sviluppare una terapia, la diffusione del virus che provoca l'AIDS continua incontrastata in buona parte del mondo. E mentre si attende che le promesse di interventi medici e clinici si concretizzino, gli sforzi per comprendere l'epidemia e fronteggiarla cambiando i comportamenti stanno avvalendosi di strumenti nuovi: modelli matematici diretti a comprendere il processo globale di trasmissione a livello di popolazione.

Nell'Africa subsahariana, e sempre più in India e in zone dell'Asia sudorientale, la velocità di diffusione dell'HIV, il virus responsabile nell'uomo della sindrome da immunodeficienza acquisita, rivela un quadro allarmante: i livelli di infezione sono estremamente elevati in gruppi quali prostitute e relativi clienti e in coloro che assumono droga per via endovenosa. Il tasso di infezione sta crescendo costantemente anche nella popolazione presa nel suo complesso.

In alcuni paesi sviluppati vi sono segnali incoraggianti di una diminuzione della velocità di crescita dell'epidemia tra gli omosessuali maschi, provocata in parte da mutamenti del comportamento. La diffusione continua però fra i tossicodipendenti e vi è una preoccupante tendenza a un aumento lento ma costante nella popolazione eterosessuale. Attualmente la diffusione dell'HIV fra gli eterosessuali in America Settentrionale e in Europa occidentale è tipicamente limitata alle minoranze povere dei centri urbani. Molti osservatori ritengono che questa tendenza rifletta semplicemente i primi stadi di un'epidemia a diffusione più vasta.

L'ottimismo nei confronti di una rapida messa a punto di vaccini è svanito. I problemi pratici ed etici che si oppongono alla sperimentazione di possibili sostanze in popolazioni umane e la necessità di un gran numero di primati per le prove di laboratorio hanno affievolito le speranze iniziali. In assenza di vaccini, l'unico modo per limitare la diffusione dell'infezione sta nell'informazione. Il pubblico deve comprendere in che modo si trasmette il virus e quali cambiamenti delle abitudini sessuali possono ridurre il rischio di infezione.

I modelli matematici potrebbero rivelarsi potenti strumenti per favorire questa consapevolezza. In epidemiologia, come del resto in altri campi della scienza, la matematica offre un metodo per trattare in maniera precisa problemi complessi. In particolare, in epidemiologia questi modelli possono aiutarci a identificare ciò che è necessario misurare e a interpretare i risultati ottenuti. Cosa ancor più importante, ci permettono di distinguere i complessi rapporti tra la biologia dell'infezione in un individuo e la trasmissione dell'infezione in una comunità.

I modelli matematici delineano una espansione lenta ma continua della epidemia di AIDS nel corso dei decenni a venire. Dato che rivelano un andamento in cui il numero di casi di infezione da HIV (e quindi di AIDS) aumenta sempre più velocemente nel tempo, come accade per il calcolo degli interessi composti, questi modelli possono svolgere un ruolo importante nel convincere Governi e organizzazioni umanitarie internazionali dell'opportunità di un intervento immediato.

Il virus dell'AIDS si è evoluto quasi certamente in Africa. Le omologie fra i ceppi che si sviluppano nell'uomo e i virus scoperti in primati selvatici nella regione subsahariana sono una buona indicazione di questa origine. Nei centri urbani africani più colpiti, dal 20 al 30 per cento delle donne gravide è sieropositiva nei confronti dell'HIV. Questo livello di infezione è stato talvolta attribuito alle condizioni socioeconomiche.

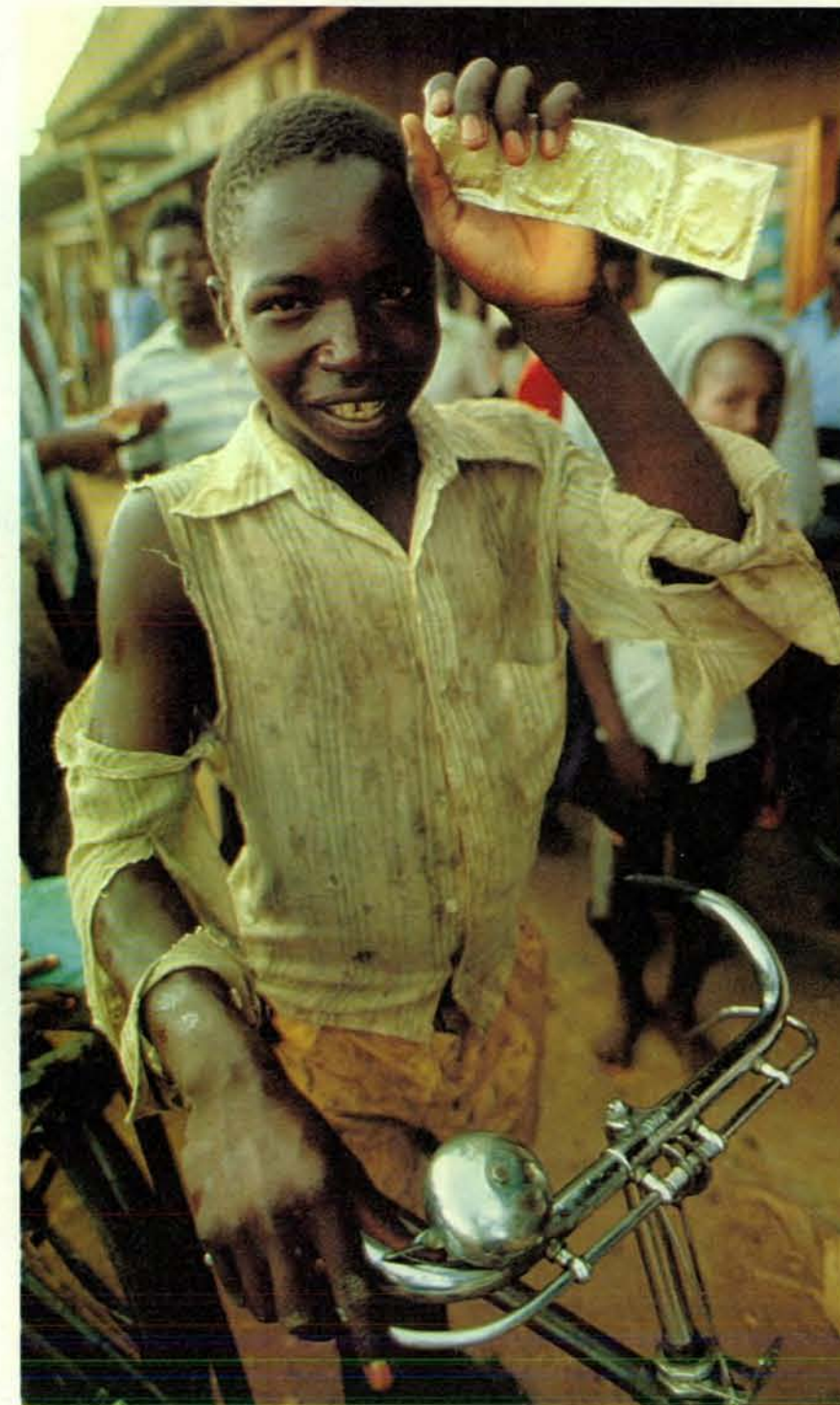
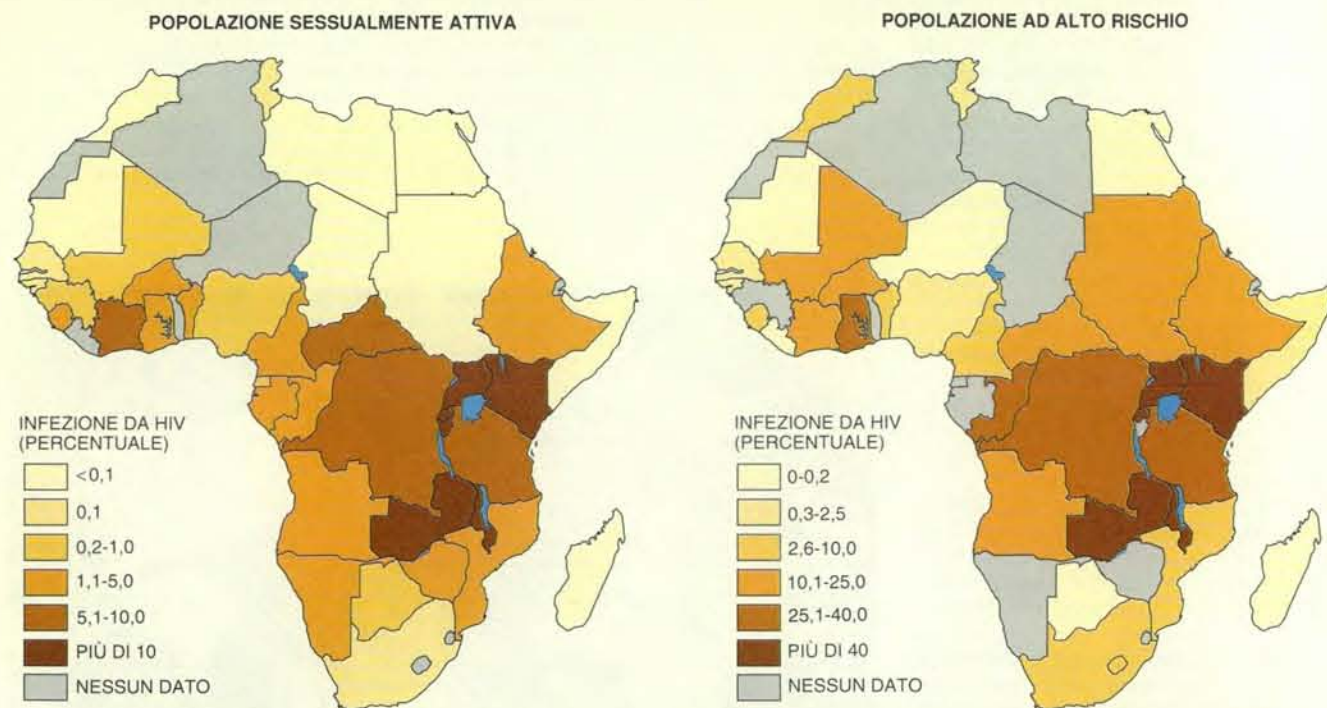
È più probabile che l'alto tasso di infezione sia una conseguenza del lungo periodo durante il quale il virus ha potuto diffondersi in queste aree: l'epidemia si trova qui semplicemente a uno stadio più avanzato. L'esame delle sequenze nucleotidiche indica che il virus umano potrebbe essersi diffuso lentamente in determinate regioni dell'Africa a partire da 100-200 anni fa, o forse anche da prima. È possibile che nel passato le complesse fasi iniziali dell'epidemia (sotto forma di episodi sporadici localizzati in aree rurali) non siano state riconosciute perché confuse nel «rumore di fondo» di altre infezioni e malattie. Oppure è possibile che nell'uomo le forme ancestrali di HIV non causassero l'AIDS.

L'epidemia si sviluppò rapidamente all'inizio degli anni ottanta fra coloro che facevano uso di droghe per via endovenosa e fra gli omosessuali maschi negli Stati Uniti e in Europa occidentale. Questo rapido avanzamento fu indubbiamente il risultato dell'introduzione del virus in comunità i cui comportamenti erano particolarmente idonei alla diffusione virale: inoculazione di droghe con scambio di siringhe e rapporti sessuali frequenti con diversi partner.

Nell'Africa subsahariana la situazione è più difficile da interpretare, ma il rapido aumento della malattia verificatosi di recente riflette probabilmente il normale sviluppo di un'epidemia. Supponendo una crescita esponenziale, con un tempo di raddoppiamento per esempio di tre anni, occorrerebbero 30 anni perché la prevalenza dell'infezione da HIV variasse dallo 0,001 per cento all'uno per cento, ma solo tre anni perché passasse dal 10 al 20 per cento. Il tempo di raddoppiamento potrebbe essere stato più lungo in passato. Diversi fattori sociali ed economici potrebbero aver contribuito ad accelerare la diffusione dell'infezione (e ad abbreviare il tempo di raddoppiamento): movimenti di popolazione provocati da conflitti, per esempio, o trasferimenti di lavoratori da un luogo a un altro, eventi che potrebbero aver fatto aumentare i contatti con prostitute a causa della separazione da mogli e fidanzate.

Quali che siano le cause della variazione nel tempo della diffusione in località diverse, si allarga sempre più la consapevolezza che, oltre a colpire gruppi minoritari, l'HIV rappresenta una grave minaccia per la maggior parte degli adulti sessualmente attivi e per i figli di madri sieropositive. I casi di AIDS registrati nel primo decennio di epidemia conclamata (gli anni ottanta) non sono che una piccola frazione di quelli che verranno registrati negli anni novanta.

La maggioranza dei casi denunciati si sposterà rapidamente dal mondo occidentale, soprattutto gli Stati Uniti, ai paesi in via di sviluppo. Considerando il lungo periodo di incubazione dell'AIDS, che negli adulti può arrivare fino a 10 anni, l'incidenza attuale della malattia (ufficialmente 350 000 casi, ma secondo alcune stime essi sarebbero oltre un milione) probabilmente riflette solo il 10 per cento circa del numero reale di persone infettate dal virus. Inoltre, anche in queste fasi iniziali della pandemia, dati recenti indicano già l'AIDS come principale causa di morte negli adulti e causa importante di mortalità infantile in alcune città dell'Africa.

La potenziale gravità della pandemia è ancora accentuata dalla straordinaria «plasticità» del genoma dell'HIV. In campioni virali isolati in tempi diversi dallo stesso paziente oppure da pazienti differenti è possibile osservare una grande diversità genetica; questa diversità è tale che i campioni virali devono essere considerati popolazioni di genomi strettamente correlati, che prendono comunemente il nome di quasispecie. La storia ha dimostrato che è estremamente difficile mettere a punto vaccini protettivi contro organismi che sono in grado di mutare rapidamente e pre-

Un ragazzo di Kinoni (Uganda) offre profilattici. La distribuzione gratuita di profilattici fa parte di un programma del Governo per limitare la diffusione dell'AIDS. Il programma è diretto soprattutto agli uomini che si trasferiscono nei centri urbani in cerca di lavoro e che hanno contatti sessuali frequenti con molte partner.

La diffusione dell'infezione da HIV fra gli individui eterosessuali nei paesi africani ha un ampio intervallo di variabilità. Sebbene l'andamento geografico del grado di diffusione dell'HIV mostri notevoli analogie tra la popolazione nel suo complesso e i gruppi ad alto rischio, i livelli di infezione sono molto più elevati per questi ultimi. I dati forniti dallo US Bureau of the Census riguardano il periodo che va dal 1985 al 1990; i livelli attuali di infezione sono indubbiamente più alti.

sentare antigeni al sistema immunitario continuamente diversi.

In passato l'aumento di diversità virale durante il periodo di incubazione è stato interpretato come conseguenza della lenta distruzione del sistema immunitario. Via via che questo viene danneggiato, un numero maggiore di tipi virali riesce a sopravvivere nell'organismo. Ma la schiera differenziata di antigeni è veramente la conseguenza della distruzione del sistema immunitario o ne è piuttosto la causa? Lavori recenti basati su semplici modelli sembrano dar credito alla seconda ipotesi.

Durante il periodo di incubazione, si osservano elevati livelli ematici del virus per un intervallo breve ma variabile (tipicamente da qualche settimana a qualche mese) successivo all'infezione. Poco dopo questo picco iniziale, è possibile rilevare la presenza di anticorpi nel sangue. È però difficile isolare il virus in seguito, durante il lungo e variabile periodo asintomatico fra l'infezione e l'insorgenza dell'AIDS. Il periodo di incubazione è caratterizzato da un basso livello di replicazione virale e da un'eventuale lenta diminuzione di un tipo di linfociti, denominati $CD4^+$, che hanno un ruolo fondamentale nella difesa organizzata del sistema immunitario contro l'HIV. In certi pazienti si osservano piccoli aumenti dei livelli virali per brevi periodi nel corso dell'incubazione. Via via che compaiono i sintomi della malattia, diventa sempre più facile isolare il virus, e nei malati di AIDS il livello di virus nel sangue è tipicamente elevato. Dopo 10 anni dall'infezione, negli omosessuali maschi circa il 50 per cento dei soggetti ha manifestato la malattia; a diagnosi avvenuta, la speranza di vita è di circa 1-2 anni in assenza di terapia con farmaci quali la azidotimidina (AZT).

Un esempio di questa tendenza, riguardante due omosessuali maschi di Amsterdam, è riportato nell'illustrazione a pagina 22. In entrambi i pazienti la diversità antigenica era inizialmente bassa, ma aumentò lentamente durante il periodo asintomatico. Uno dei pazienti manifestò l'AIDS e nel suo caso si può notare che la diversità virale culminò prima della comparsa della malattia conclamata e in seguito andò declinando. Questa diminuzione potrebbe apparire a prima vista contraria alla logica, ma è esattamente ciò che la teoria prevede. Durante il lungo periodo di incubazione si ha un aumento del numero di ceppi di HIV, ma tutti sono soggetti al controllo del sistema immunitario. Pertanto, via via che il numero dei ceppi aumenta, i membri di ognuno di essi sono presenti in numero pressoché uguale. L'aumento di diversità è quindi rivelabile in campioni relativamente piccoli di siero o cellule. Ma nella fase di AIDS conclamata, quando i livelli virali aumentano e il sistema immunitario crolla, quei ceppi di HIV che hanno un'alta velocità di replicazione si avvantaggiano e tendono a essere i soli presenti nei campioni.

Un modello matematico di questo processo fornisce un andamento della variazione temporale del livello e della diversità del virus che corrisponde molto bene alla tendenza osservata. Il modello prevede che la diversità virale abbia una soglia critica sotto la quale il sistema immunitario può limitare l'abbondanza del virus e sopra la quale la crescita virale non può più essere contrastata. In altri termini, il modello indica che lo sviluppo di una grave immunodeficienza nei pazienti infettati da HIV è una conseguenza della capacità del virus di produrre continuamente proprie varianti e di uccidere i linfociti $CD4^+$. La diversità antigenica è la causa, non la conseguenza, dell'immunodeficienza.

Il modello può essere utilizzato anche per verificare l'efficacia del trattamento con AZT. Sembra che questa terapia ostacoli il progredire dell'AIDS. Differenti tempi di inizio del trattamento durante il lungo periodo di incubazione dell'AIDS indicano che un intervento il più possibile precoce può prolungare in maniera significativa l'intervallo che precede la comparsa di una grave immunodeficienza.

I modelli matematici stanno anche chiarendo la dinamica della trasmissione dell'HIV all'interno di varie comunità e fra l'una e l'altra. La potenzialità di diffusione dipende dall'entità del «tasso riproduttivo» dell'infezione. Questo valore, indicato con R_0, definisce il numero medio di casi secondari di infezione generati da un caso primario in una popolazione in cui quasi tutti i membri siano suscettibili all'infezione, come accade nelle prime fasi di un'epidemia. Se R_0 è maggiore di 1, ciascun caso di infezione ha più di un caso «figlio» e il risultato è una reazione a catena, ossia un'epidemia. Se R_0 è minore di 1, l'infezione non può mantenersi. Più è grande il valore di R_0, più è breve il tempo che occorre perché il numero di casi raddoppi.

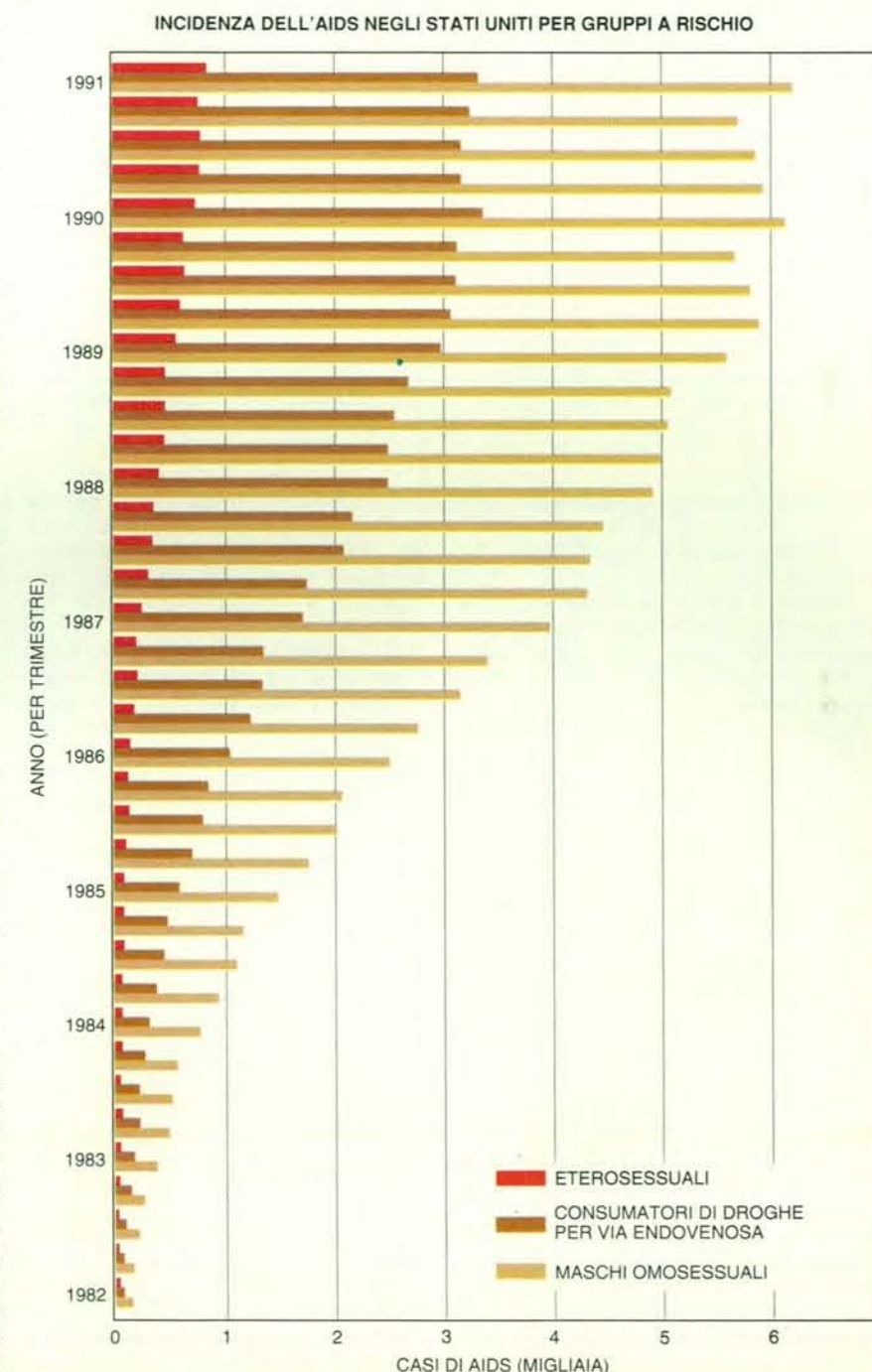
Per quanto riguarda l'HIV, R_0 è essenzialmente il prodotto di tre fattori: la probabilità media che una persona sieropositiva infetti un partner nel corso della relazione, il numero medio di partner avuti nell'unità di tempo e la durata media dell'infettività. Sebbene questa semplice definizione nasconda numerose complicazioni, è utile per concentrare l'attenzione su ciò che si deve misurare per comprendere la velocità di diffusione dell'infezione.

Due complicazioni associate alla definizione di R_0 sono che il suo valore differisce in misura significativa da un gruppo a rischio a un altro e che esso può variare nel tempo in risposta a cambiamenti di comportamento. La variabilità dell'attività sessuale nell'ambito di una data popolazione è di particolare importanza nel determinare la diffusione dell'infezione. Qual è il tipo e quale la frequenza dell'attività sessuale? Quanto è frequente il cambiamento di partner? Qual è la rete di contatti sessuali? Purtroppo sono domande a cui è difficile rispondere, in parte a causa della riluttanza di Governi e cittadini nei riguardi dello studio di quello che, in fin dei conti, è un comportamento personale.

Gli scarsi dati disponibili sulla frequenza dei cambiamenti di partner e dei rapporti sessuali rivelano una grande eterogeneità in seno alle varie popolazioni. All'interno di questa variabilità si possono distinguere andamenti precisi, che sembrano essere abbastanza costanti sia da comunità a comunità (eterosessuali e omosessuali, per esempio) sia da cultura a cultura (per esempio, paesi sviluppati e in via di sviluppo).

Il concetto di variabilità del comportamento sessuale può essere incluso nella definizione di R_0 sostituendo il numero medio di partner acquisiti nell'unità di tempo con un'espressione che tenga conto sia della media sia dello scarto rispetto alla media. Uno scarto elevato aumenta il valore di R_0. Logicamente, gli individui che cambiano spesso partner hanno una probabilità maggiore di ricevere e di trasmettere l'infezione. I dati statistici dimostrano che questi individui costituiscono una piccola frazione della popolazione totale. A meno del 20-30 per cento degli individui studiati va attribuito il 70 per cento delle relazioni sessuali registrate. Paradossalmente, allora, mentre uno scarto elevato aumenta il successo globale della trasmissione dell'HIV (in termini di R_0) e quindi accresce la probabilità che si abbia un'epidemia, d'altra parte implica anche che l'infezione resti limitata, fino a un certo punto, alla piccola frazione di individui sessualmente molto attivi.

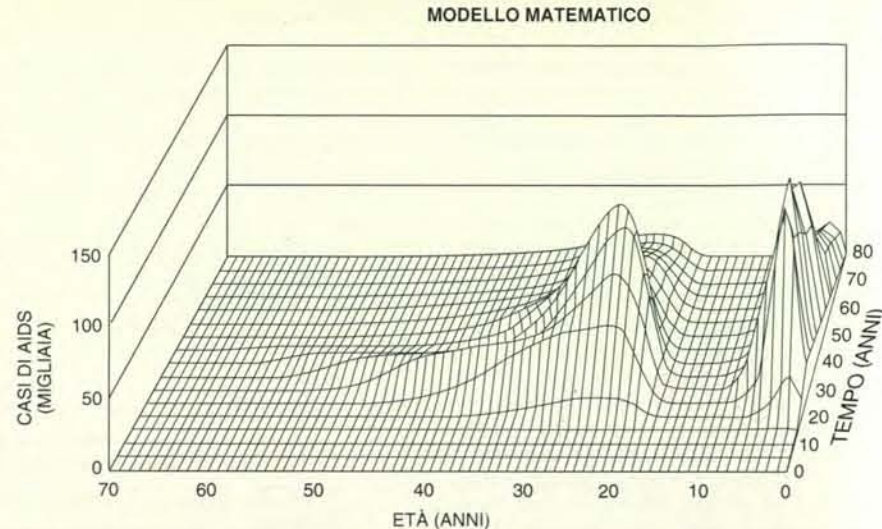
Il grado di contenimento dell'infezione in un piccolo nucleo di persone sessualmente attive dipende dalla rete di contatti sessuali fra individui che sono molto attivi e individui che non lo sono. Per definire questa rete occorre ottenere informazioni non solo sul numero di partner, ma anche sulla loro identità. Ciò è necessario per stabilire quanto spesso costoro cambino a loro volta partner. La comprensibile riluttanza degli intervistati a fornire l'identità dei partner rende questi studi molto difficili.

In assenza di dati, non abbiamo altra scelta che affidarci di nuovo a modelli matematici per valutare quanto sia importante acquisire queste informazioni. I modelli possono indicare il grado in cui i rapporti fra persone dal comportamen-

Il numero dei casi di AIDS registrati negli Stati Uniti differisce a seconda del gruppo a rischio. Per i tre gruppi che sono qui considerati, i dati, forniti dalla HIV/AIDS Division del Center for Infectious Diseases, sono aggiornati al giugno 1991.

to simile influiscono sull'epidemia. In questo caso non basta più definire un singolo valore di R_0. Dobbiamo costruire una matrice di valori per caratterizzare la trasmissione dell'infezione all'interno dei vari gruppi e fra l'uno e l'altro gruppo. A seconda del tipo di mescolamento dei gruppi, alcuni elementi della matrice potranno avere un valore maggiore di uno (l'infezione si diffonde in quel gruppo) mentre altri lo avranno minore di uno (l'infezione si estingue). In parole semplici, alcuni gruppi possono mantenere nella popolazione complessiva un'infezione che, in loro assenza, scomparirebbe. Questa osservazione indica che le campagne informative dovrebbero essere rivolte agli individui che più contribuiscono alla proliferazione dell'infezione.

L'epidemia prospettata nello scenario in cui si ha un mescolamento casuale tra gruppi dal comportamento diverso è molto più grave di quella prevista nel caso precedente. All'inizio l'infezione si diffonde molto più rapidamente nel primo caso: il piccolo nucleo di individui sessualmente attivi è infettato, e si sviluppa un'epidemia ristretta, che si diffonde in misura limitata nel gruppo, molto più ampio, di individui meno attivi. Il punto su cui è bene fare attenzione è che il mescolamento fra gruppi dal comportamento simile può generare un'epidemia con un andamento a più picchi, via via che l'infezione si diffonde dai gruppi a rischio elevato a quelli a rischio minore.

Principi simili si applicano se consideriamo divisioni a scala maggiore. È probabile che in ogni dato paese l'epidemia si sviluppi in una serie di onde che passano dai gruppi a rischio elevato a quelli a rischio basso tramite contatti reciproci. Una di queste catene di contatti è costituita da uomini omosessuali, uomini bisessuali e donne eterosessuali. Un'altra catena collega i tossicodipendenti eterosessuali e i non tossicodipendenti eterosessuali. Una terza è formata dalle prostitute, dai loro clienti e dalle partner di questi ultimi. Che l'epidemia globale si manifesti come una serie di picchi distinti o come un unico picco regolare dipende sia dalla dimensione di ciascun gruppo a rischio sia, in modo più significativo, dall'entità dei contatti fra i vari gruppi.

In molti paesi occidentali sembra che il picco dei casi di infezione da HIV negli omosessuali maschi sia già stato raggiunto. Stiamo ora osservando il progredire di una seconda epidemia, che non ha ancora raggiunto il culmine, fra i tossicodipendenti che si iniettano droghe. Una terza ondata, che riguarda gli eterosessuali, è appena all'inizio, ma la sua ampiezza complessiva è ancora incerta. Nei paesi in via di sviluppo, soprattutto nell'Africa subsahariana, il picco iniziale verificatosi fra le prostitute viene ora superato da un'infezione molto più disseminata nella popolazione complessiva

Gli effetti demografici dell'AIDS, studiati matematicamente, vengono qui confrontati con quelli osservati in Uganda. L'impatto previsto da un modello di trasmissione dell'HIV strutturato per classi di età mostra fondamentali somiglianze con la distribuzione per età e sesso dei casi di AIDS accertati nel 1988 in Uganda. In particolare, sia i risultati del modello sia i dati reali mostrano due picchi di incidenza: uno corrispondente alla prima infanzia e uno al decennio sessualmente più attivo.

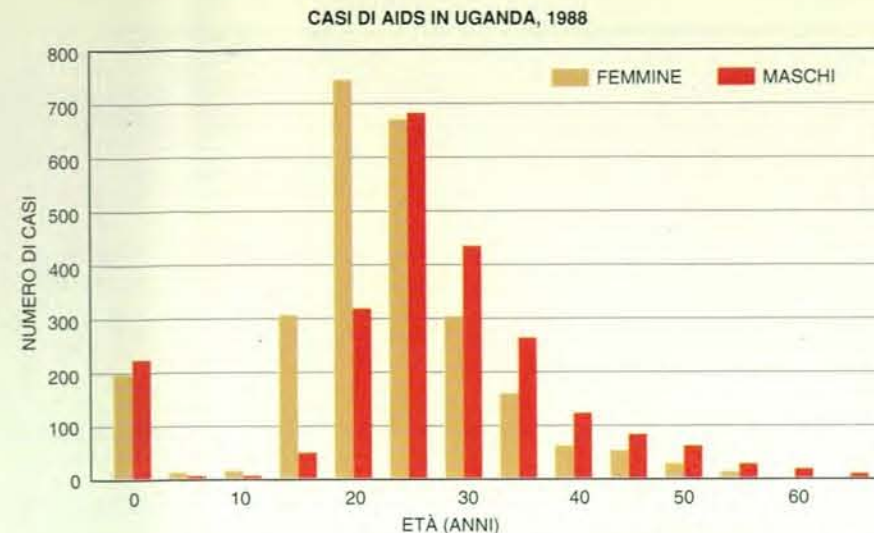
eterosessuale. Nell'Asia sudorientale si comincia a distinguere un andamento simile, sebbene in Thailandia si sia osservata la prima ondata fra i tossicodipendenti oltre che fra le prostitute. Sia in Africa sia in Asia l'infezione da HIV è sufficientemente diffusa nella popolazione da creare proccupazione per le conseguenze demografiche ed economiche che si manifesteranno nei prossimi decenni.

Per valutare l'impatto demografico dell'AIDS nelle regioni del mondo più colpite, dobbiamo mettere a punto modelli che combinino descrizioni della diffusione dell'HIV e della crescita della popolazione. Le prime analisi, nel 1988, indicavano che l'AIDS sarebbe in grado di invertire una crescita demografica positiva in una scala di tempo di alcuni decenni. Anche la crescita annuale del 3 per cento o più che si ha nell'Africa subsahariana potrebbe essere trasformata in un andamento negativo. Queste prime previsioni sono state oggetto di molte controversie, soprattutto a causa dell'incertezza nell'assegnazione dei parametri e nella mancata considerazione delle differenze di attività sessuale nelle varie comunità e da una comunità a un'altra.

Più di recente si sono accumulati dati che aiutano a stimare i parametri epidemiologici più importanti, ma questi tentativi sono affetti ancora da gravi incertezze. Da studi condotti nei paesi sviluppati, sembra che il periodo di incubazione medio dell'AIDS per i casi di trasmissione del virus dalla madre al feto sia di uno o due anni. A quanto pare, dal 13 al 30 per cento dei bambini nati da madri sieropositive nei paesi sviluppati manifesta l'infezione; in Africa questo valore è più vicino al 40 per cento.

L'incertezza più grande nel valutare la velocità di trasmissione del virus riguarda l'andamento più consueto del comportamento sessuale nelle varie comunità e da una comunità all'altra. L'Organizzazione mondiale della sanità ha dato il via ad ampie indagini sul comportamento sessuale in Africa, e i dati cominciano ad affluire. Ma per ora le nostre limitate conoscenze sui processi che regolano il comportamento sessuale in società diverse sono per lo più di natura aneddotica.

Date le circostanze, i modelli matematici possono farci ancora da guida, questa volta per identificare i collegamenti fra l'attività sessuale e l'andamento osservato della diffusione dell'HIV. Negli ultimi anni sono stati sviluppati modelli che affrontano diversi aspetti di questi problemi, anche se bisogna sottolineare che gli sviluppi teorici hanno notevolmente sopravanzato i dati disponibili. Il ruolo principale di questi modelli è di fornire indicazioni qualitative riguardo all'interpretazione degli andamenti osservati. Il modello dovrebbe anche fornire un quadro entro cui valutare l'impatto potenziale di interventi diversi. Le nostre conoscenze possono svilupparsi meglio aggiungendo complessità a poco a poco, in maniera analoga a quella seguita da uno sperimentatore che cambia un fattore per volta, mantenendo costanti tutti gli altri.

Come illustrazione di ciò che possono dirci questi modelli, consideriamo due esempi degli effetti della variabilità dei contatti sessuali sul potenziale esito dell'AIDS in Africa. Il nostro primo esempio riguarda l'influenza di vari tipi di contatti fra diverse classi di età dei due sessi. In particolare ci concentreremo sulla tendenza degli uomini africani ad avere relazioni sessuali con donne da 5 a 10 anni più giovani di loro. A parità

di altri fattori, i modelli che tengono conto della preferenza di età osservata fanno pensare a un impatto demografico notevolmente più grave rispetto alle previsioni basate su contatti limitati alla stessa classe di età. Il maggior impatto è una conseguenza della concentrazione dell'infezione in giovani donne che stanno facendo il loro ingresso nell'età feconda. I modelli indicano anche che il rapporto fra uomini e donne infettati dall'HIV è destinato a cambiare nel corso dell'epidemia. Questa variazione potrebbe spiegare perché in Africa siano stati registrati valori ampiamente discordanti, con un rapporto uomini-donne che va da 1:1 a 1:2 circa.

Il nostro secondo esempio riguarda il tipo dei rapporti che intercorrono fra uomini e donne che cambiano partner con frequenza più o meno elevata. Il modello più semplice - basato sulla stratificazione della popolazione per sesso e classe di attività sessuale - illustra bene il significato delle reti di contatti sessuali.

Quando uomini sessualmente molto attivi (come i lavoratori immigrati nei centri urbani) hanno contatti rilevanti con donne molto attive (per esempio prostitute) e contatti poco frequenti con donne meno attive (mogli o fidanzate), può svilupparsi un'epidemia multipla. Dapprima si ha un'epidemia a rapido sviluppo nella piccola frazione di uomini e donne molto attivi. All'episodio iniziale fa seguito un'epidemia a sviluppo più lento, ma molto più vasta, che coinvolge gli uomini e le donne a bassa attività sessuale (la maggioranza della popolazione). L'epidemia nelle classi a elevata attività funge da «seme» per quella a crescita più lenta, e i picchi delle due classi possono essere separati da un intervallo di 10 anni o più.

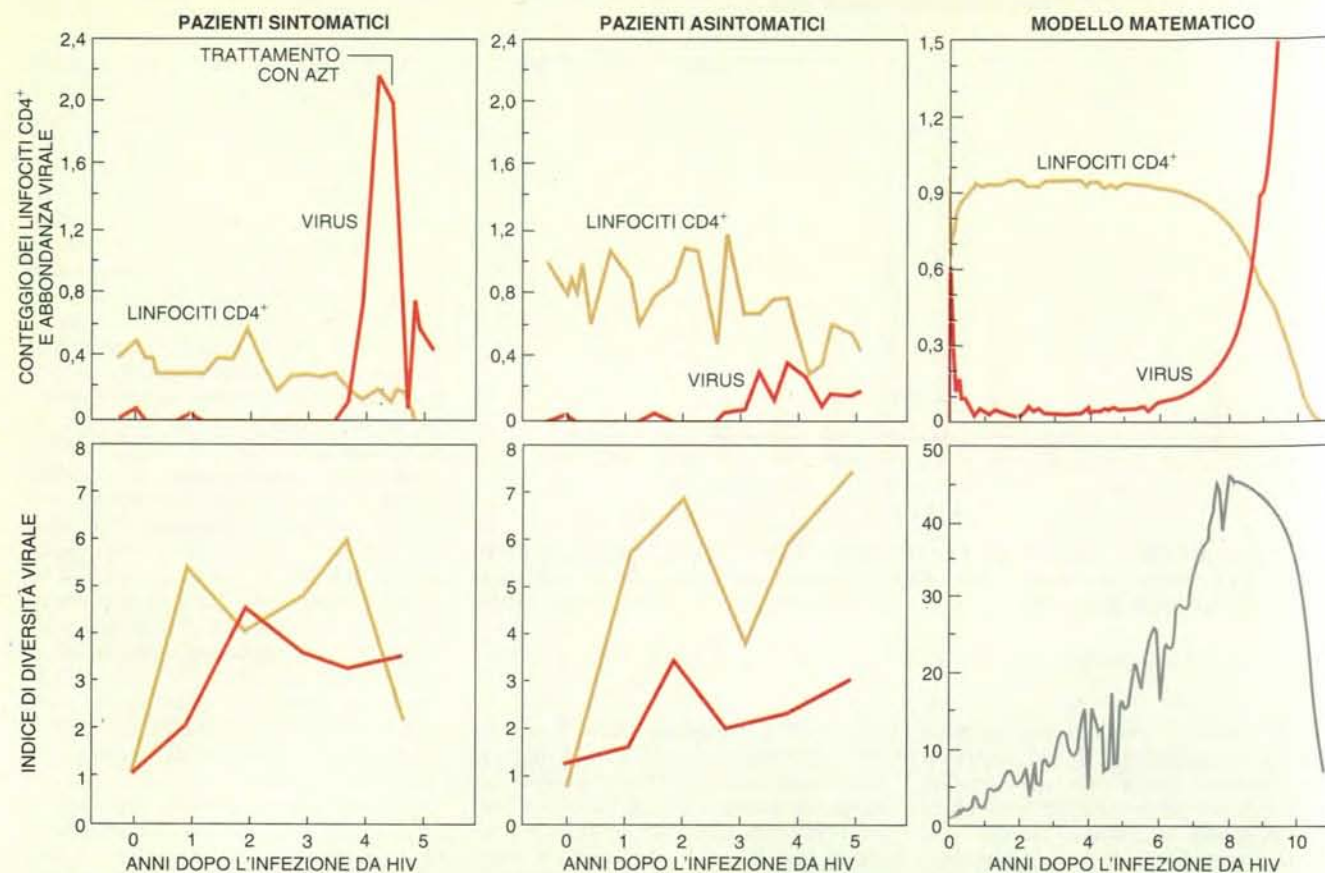
Lo scenario previsto da questo semplice modello potrebbe riflettere ciò che sta effettivamente avvenendo in città come Nairobi. Qui i livelli di infezione sono elevati nelle prostitute (dal 60 all'80 per cento), da moderati ad alti nei loro clienti (dal 20 al 40 per cento) e bassi nelle donne gravide della popolazione complessiva (dal 5 al 6 per cento). Il modello indica che i bassi livelli della popolazione nel suo insieme aumenteranno nel prossimo decennio. Questo aumento darà inizio a una seconda e molto più vasta epidemia, analogamente a quanto è già accaduto nelle città del Malawi, della Tanzania e dell'Uganda.

I dati certi disponibili indicano che, in assenza di grandi cambiamenti nel comportamento o di una messa a punto di farmaci più efficaci, è probabile che nei prossimi decenni l'AIDS provochi serie alterazioni demografiche in alcuni paesi africani. Sembra sempre più probabile che questo andamento sia destinato a ripetersi in certe zone dell'India e dell'Asia sudorientale.

Che cosa si può fare per arginare il diffondersi dell'infezione? I modelli che cercano di valutare l'influenza di un cambiamento del comportamento sulla velocità di diffusione dell'infezione mostrano quale importanza abbia il momento in cui il cambiamento si verifica. Gli effetti di tutto ciò non sono intuitivamente ovvi, dato che l'epidemia ha carattere non lineare. Cambiamenti introdotti in una fase precoce dell'epidemia hanno un effetto infinitamente più rilevante di cambiamenti analoghi introdotti in seguito. Di conseguenza si dovrebbero destinare risorse significative allo scopo di indurre questi mutamenti, nel tentativo di prevenire un'epidemia mortale che potrebbe essere ampiamente diffusa fra 10 o 20 anni.

Campagne di informazione e la distribuzione di profilattici mirate ai gruppi ad alto rischio saranno sempre misure utili nelle fasi precoci dell'epidemia, quando l'infezione nella popolazione complessiva è limitata. Questa politica può essere molto vantaggiosa in paesi, come la Nigeria, dove i livelli di infezione da HIV in gruppi ad alto rischio quali le prostitute e i loro clienti sono da bassi a moderati, e molto bassi nelle donne in gravidanza. Se i contatti sessuali ricadono soprattutto nell'ambito di gruppi con attività simile, un approccio diretto specificamente ai gruppi ad alto rischio è particolarmente valido. Potrebbe anche accadere che la frequenza di cambiamento di partner sessuale nella popolazione eterosessuale sia insufficiente per mantenere la trasmissione dell'HIV. Se tuttavia l'infezione ha preso piede in modo significativo in gruppi a rischio più basso, come è accaduto in Malawi, Tanzania e Uganda, l'informazione e la distribuzione di profilattici devono avere un raggio più vasto. In effetti, in queste circostanze, i modelli indicano che sarebbe particolarmente utile concentrarsi sui giovani all'inizio dell'adolescenza, prima che divengano sessualmente attivi.

Le sottili interconnessioni tra frequenza di cambiamento di partner, andamento dei rapporti tra gruppi a diversa attività sessuale e necessità di bilanciare domanda e offerta di partner nella popolazione costituiscono un esempio interessante. Supponiamo che una comunità abbia ampie vedute riguardo alla formazione di relazioni sessuali. La maggior parte degli uomini e delle donne avrà un certo numero di partner differenti ogni anno, con il risultato che gli uomini avranno contatti limitati con prostitute. Se l'informazione riduce il tasso di cambiamento di partner fra le donne, il che significa che gli uomini avranno contatti più frequenti con prostitute, il risultato potrebbe essere un'accelerazione a breve termine della diffusione del virus nella popolazione complessiva. A lungo termine, però, la dimensione totale dell'epidemia sarebbe ridotta. L'esempio mostra la necessità di valutare gli effetti dell'informazione non solo sul tasso di cambiamento di partner, ma anche sul tipo di relazioni sessuali all'interno di una data comunità.

Più in generale, un numero sempre maggiore di dati mostra l'importanza di altre malattie a trasmissione sessuale. Queste ultime sembrano aumentare la probabilità di trasmissione dell'HIV, presumibilmente a causa di lesioni nella regione genitale. Questi dati mettono in evidenza la necessità di controllare la diffusione delle malattie a trasmissione sessuale nei paesi in via di sviluppo, soprattutto in Africa. Semplici modelli che descrivono la trasmissione concomitante delle malattie veneree e dell'HIV chiariscono la necessità di questo impegno. Un altro vantaggio di un più stretto con-

Il decorso dell'AIDS in due individui omosessuali di sesso maschile di Amsterdam (*grafici a sinistra e al centro*) è riprodotto molto bene dai risultati di un semplice modello matematico (*grafici a destra*). Nella fila in alto è mostrata l'abbondanza sia del virus sia dei linfociti CD4+. La fila in basso indica che la diversità genetica del virus cresce generalmente nel tempo. Le curve in basso a sinistra e al centro si riferiscono a due indici impiegati per misurare la diversità virale.

trollo delle malattie a trasmissione sessuale è che questi programmi facilitano l'assistenza e la distribuzione di profilattici al segmento ad alto rischio della popolazione.

I paesi sviluppati hanno un'altra possibilità per ridurre al minimo gli effetti dell'AIDS: l'uso di azidotimidina e di altri farmaci che sembrano rallentare il decorso della malattia. Una sperimentazione clinica a distribuzione casuale effettuata negli Stati Uniti ha permesso di valutare l'effetto di due dosaggi differenti di AZT (500 e 1500 milligrammi al giorno) e di un placebo su pazienti asintomatici. La prova è stata interrotta prima del termine previsto perché i pazienti a cui veniva somministrata azidotimidina mostravano un considerevole rallentamento del decorso dell'AIDS e del complesso avanzato correlato all'AIDS (ARC).

Il trattamento con AZT può avere in certi casi risvolti negativi. Sebbene sia vantaggioso per l'individuo, in quanto ne prolunga la sopravvivenza, esistono situazioni in cui può essere dannoso per la comunità. A seconda della riduzione dell'infettività dovuta al trattamento, che per ora non è nota con precisione, può darsi che l'allungamento dell'incubazione e del periodo sintomatico negli individui così trattati possa aumentare l'incidenza dell'HIV nella comunità. Al limite, ciò potrebbe produrre nel complesso un incremento del numero di morti collegate all'AIDS. Dal momento che l'AIDS è una malattia mortale e che l'AZT sembra rallentarne il decorso, non sarebbe certamente etico rifiutare il trattamento ai pazienti. Tuttavia i modelli matematici confermano la possibilità che ciò che è vantaggioso per il singolo non lo debba necessariamente essere per la comunità.

È chiaro che sono necessarie ricerche virologiche di natura quantitativa per misurare l'efficacia del trattamento con farmaci sull'infettività, e perché un'infettività ridotta divenga parte dei protocolli per la messa a punto di farmaci. Inoltre queste considerazioni mettono in evidenza quanto sia desiderabile collegare il trattamento con campagne di sensibilizzazione che promuovano un'attività sessuale prudente negli individui sieropositivi.

In assenza di farmaci e vaccini efficaci, un cambiamento del comportamento sessuale rappresenta l'unica arma contro l'AIDS. I modelli matematici, con le loro indicazioni qualche volta sorprendenti e poco intuitive, possono incanalare questo cambiamento nelle direzioni più proficue.

BIBLIOGRAFIA

CHIN JAMES, *Current and Future Dimensions of the HIV/AIDS Pandemic in Women and Children* in «Lancet», 336, n. 8709, 28 luglio 1990.

NOWACK M. A., MAY ROBERT M. e ANDERSON ROY M., *The Evolutionary Dynamics of HIV-1 Quasispecies and the Development of Immunodeficiency Disease* in «AIDS», 4, n. 11, novembre 1990.

ANDERSON ROY M. e MAY ROBERT M., *Infections Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.

ANDERSON ROY M., MAY ROBERT M., BOILY M. C., GARNETT G. P. e ROWLEY J. T., *The Spread of HIV-1 in Africa: Sexual Contact Patterns and Predicted Demographic Impact of AIDS* in «Nature», 352, n. 6336, 15 agosto 1991.

Le funzioni biologiche dell'ossido di azoto

Questo composto chimico notoriamente tossico sta rivelandosi dotato di un'importanza fisiologica finora insospettata e potrebbe rappresentare il primo esempio di una classe del tutto nuova di neurotrasmettitori

di Solomon H. Snyder e David S. Bredt

Piccolo, semplice e tossico, l'ossido di azoto sembra un candidato ben poco probabile al ruolo di «tuttofare» biologico. In effetti, le funzioni dell'organismo sono in gran parte regolate da proteine e altri composti eccezionalmente grandi e complessi, quali il testosterone, che è un ormone, e l'interferone gamma, una proteina del sistema immunitario.

Questa complessità chimica sembra quasi esagerata quando la si confronta con la semplicità e la modestia dell'ossido di azoto. Questo composto, di formula NO, è gassoso a pressione atmosferica e non va confuso con il protossido di azoto (N_2O), il gas esilarante impiegato in anestesia. L'ossido di azoto è notoriamente tossico data la sua struttura da radicale libero: l'elettrone in più disponibile per il legame lo rende assai reattivo. E benché si sappia da molto tempo che i batteri contengono ossido di azoto, nessuno aveva previsto che un agente così reattivo potesse avere una funzione importante nei mammiferi.

Questa convinzione fu smentita nel 1987, quando una serie di scoperte avvenute nel corso di ricerche con indirizzo e scopi diversi contribuì a comporre un quadro coerente che rivelava le importanti funzioni biologiche dell'ossido di azoto. Questa sostanza è forse uno dei più importanti messaggeri chimici: consente ai leucociti di distruggere cellule neoplastiche e batteri, e ai neurotrasmettitori di dilatare i vasi sanguigni.

Inoltre l'ossido di azoto, proprio come un neurotrasmettitore, funge da messaggero per i neuroni; anzi, potrebbe rivelarsi il primo di una classe di neurotrasmettitori differenti da quelli già noti. Comprendere i meccanismi molecolari di azione di questo composto, la sua distribuzione e le sue correlazioni con altre sostanze attive nell'organismo potrebbe essere illuminante sia per le ricerche sulla memoria sia per il trattamento di certe patologie degenerative del sistema nervoso.

Dai primi studi sull'ossido di azoto pareva che questo composto non potesse avere che effetti nocivi. Estremamente instabile, può esistere per non più di 6-10 secondi prima di essere convertito da ossigeno e acqua in nitrati e nitriti. Benché nell'uomo si abbia escrezione di nitrati, si è sempre ritenuto che questi composti derivassero solo dagli alimenti ingeriti. Quando, nel 1956, P. N. Magee e J. M. Barnes dei Medical Research Council Laboratories nel Surrey (Inghilterra) riferirono che l'organismo trasforma i nitrati presenti nei cibi conservati in nitrosammine - che hanno azione cancerogena - molti ridussero il consumo di salumi e di altri alimenti ricchi di nitrati.

Nel 1981 Steven R. Tannenbaum e collaboratori del Massachusetts Institute of Technology osservarono che notevoli quantità di nitrati venivano escrete anche da esseri umani e ratti sottoposti a una dieta povera di nitrati. Ovviamente la dieta non era l'unica fonte di questi composti. Qual era dunque la loro origine? Tannenbaum trovò un indizio importante osservando che uno dei soggetti in esame, un uomo colpito da diarrea infettiva, eliminava grandi quantità di nitrati con le urine. I processi infiammatori associati alla diarrea erano evidentemente responsabili della formazione dei nitrati. Tannenbaum notò anche che l'inoculazione nei ratti di endotossina batterica, che induce una risposta infiammatoria ai batteri, stimolava l'escrezione di nitrati.

La causa esatta della formazione dei nitrati e la sua connessione con una risposta infiammatoria sono state infine individuate da Michael A. Marletta e Dennis J. Stuehr dell'Università del Michigan e da John B. Hibbs, Jr., dell'Università dello Utah. Marletta, già allievo di Tannenbaum al MIT, era interessato al ruolo svolto dal sistema immunitario nella produzione di nitrati stimolata dalle endotossine. Egli scoprì che i topi con una carenza genetica di macrofagi espellevano quantità ridotte di nitrati e individuò in questo modo un legame fra la presenza dei macrofagi e quella dei nitrati.

Marletta proseguì le proprie indagini preparando colture di macrofagi e addizionando al mezzo di coltura endotossina batterica e interferone gamma, una proteina modulatrice del sistema immunitario che attiva altre cellule immunitarie ed è secreta dai linfociti *T*. In seguito a questo trattamento, i macrofagi acquisirono improvvisamente la capacità di produrre nitrati.

Sottoponendo le colture a diversi tipi di analisi specifiche, Marletta scoprì anche che i macrofagi non erano in grado di produrre nitrati in assenza di arginina, un amminoacido normalmente contenuto nel mezzo di coltura. Questa scoperta gli permise di dimostrare che uno specifico enzima presente nei macrofagi trasforma l'arginina in un composto chimico intermedio; quest'ultimo risultò essere l'ossido di azoto, che viene rapidamente convertito in nitriti e nitrati.

Nel frattempo Hibbs, indipendentemente, valutava la capacità dei macrofagi di distruggere cellule neoplastiche e batteri. Dopo aver messo nella stessa coltura cellule tumorali e macrofagi, osservò che questi ultimi perdevano la capacità di eliminare le cellule se si escludeva l'arginina dal mezzo di coltura. Hibbs dimostrò inoltre che l'arginina veniva trasformata non solo in nitrati, ma anche in un altro amminoacido, la citrullina; in questo modo riuscì a provare che era un enzima specifico a produrre ossido di azoto dall'arginina.

Proseguendo le ricerche Hibbs dimostrò che l'ossido di azoto è tossico per le cellule neoplastiche quanto i macrofagi attivati; individuò inoltre il primo inibitore dell'enzima che sintetizza questo gas evidenziando il fatto che un derivato dell'arginina - specificamente una sua forma metilata - impediva sia la formazione di nitrati sia la distruzione delle cellule neoplastiche da parte dei macrofagi. In assenza di un enzima che produca ossido di azoto dall'arginina, i macrofagi non possono svolgere la loro funzione difensiva.

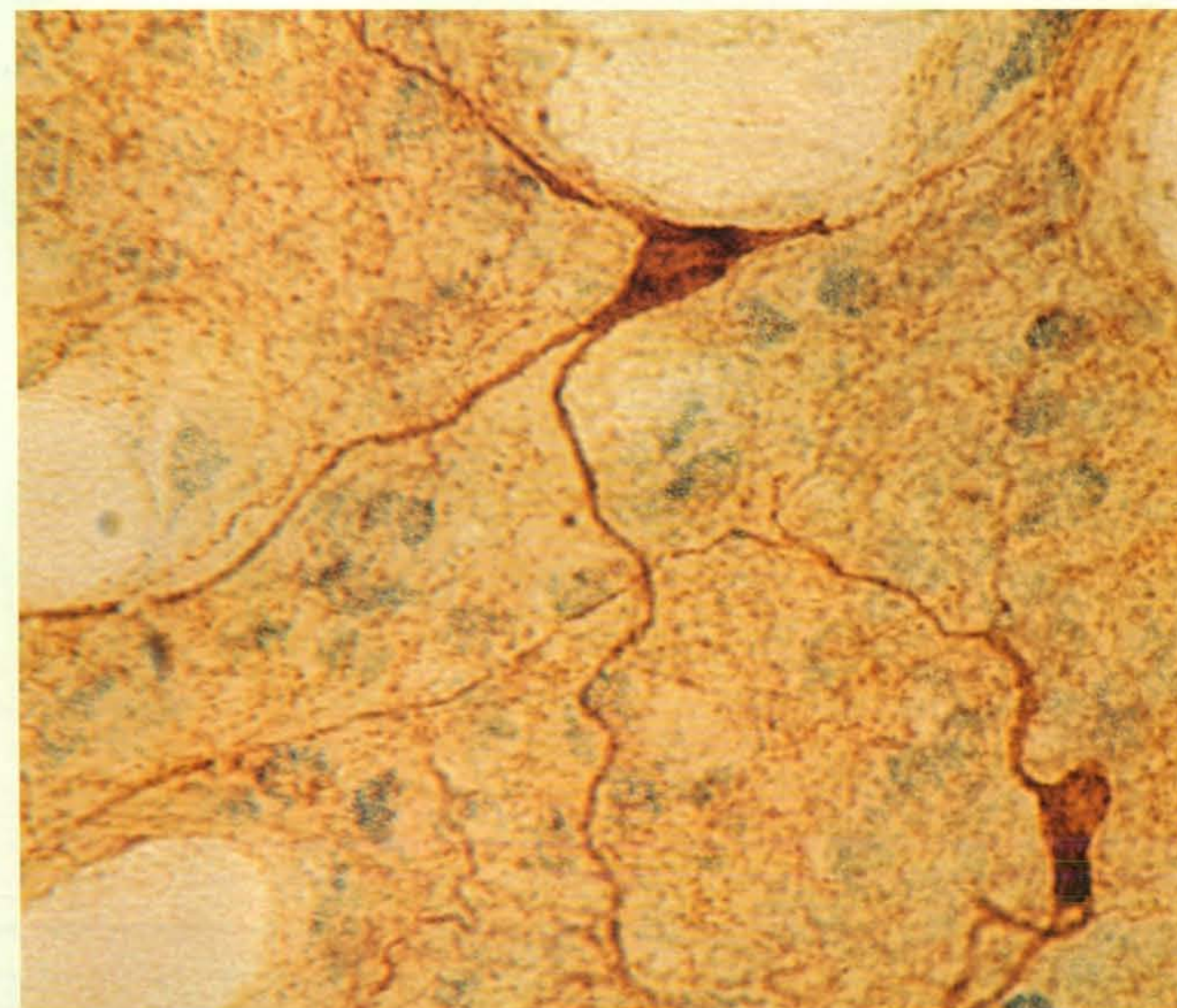
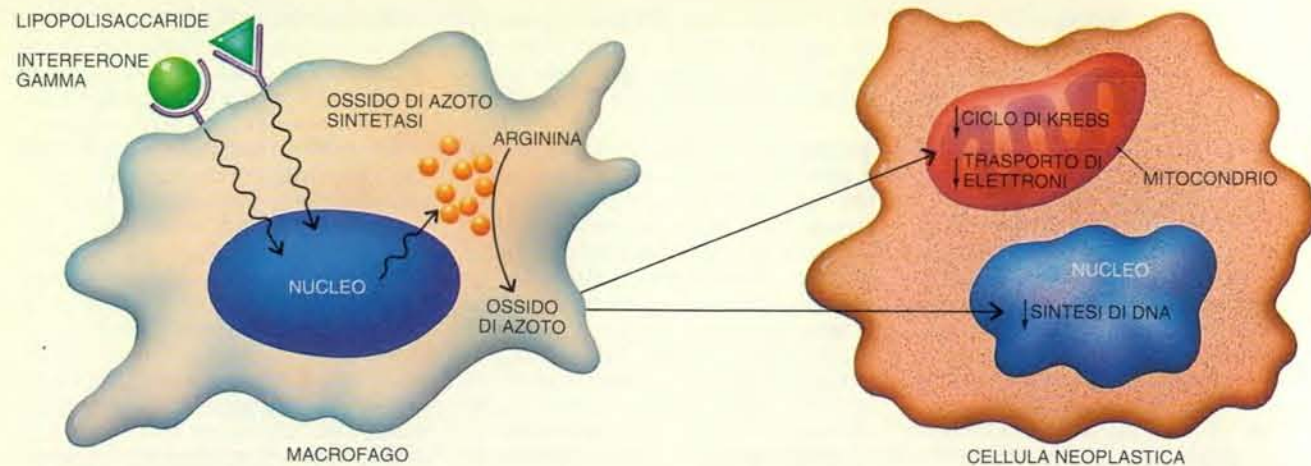
Questo è il punto a cui erano giunte le ricerche alcuni anni fa. Quando i macrofagi sono attivati dalle endotossine o dai linfociti *T*, reagiscono trasformando l'arginina in ossido di azoto, e a sua volta questo composto tossico e reattivo consente loro di distruggere batteri, funghi e cellule neoplastiche.

In una serie di indagini del tutto indipendenti, altri ricercatori hanno stabilito che l'ossido di azoto può agire come messaggero chimico. In questo ruolo esso manifesta due tipi di attività: il primo riguarda i meccanismi di dilatazione dei vasi sanguigni a opera dei neurotrasmettitori; il secondo si ricollega ai farmaci che alleviano i sintomi dell'angina pectoris, una patologia cardiovascolare caratterizzata dalla costrizione delle arterie coronarie. Le ricerche su questi due tipi di attività sono di recente confluite, rivelando tutta la complessità delle funzioni biologiche dell'ossido di azoto.

I vasi sanguigni vengono dilatati da neurotrasmettitori, come l'acetilcolina, che provocano il rilassamento dello strato muscolare della parete. Altri neurotrasmettitori come la noradrenalina contrastano questo effetto facendo contrarre lo strato muscolare e restringendo il lume del vaso. Dato che i recettori per la noradrenalina sono situati sulle cellule muscolari, si riteneva che queste avessero anche recettori per l'acetilcolina.

Nel 1980 si dimostrò che ciò non era vero. Robert F. Furchgott, uno dei massimi esperti di farmacologia cardiovascolare che lavorava al Downstate Medical Center di Brooklyn, notò che il rilassamento dei vasi sanguigni prodotto dall'acetilcolina veniva meno se si asportava dalla parete del vaso il tessuto endoteliale. (L'endotelio è un sottile strato di cellule che riveste la superficie interna dei vasi sanguigni, immediatamente sopra allo strato muscolare.) Egli dimostrò sperimentalmente che l'acetilcolina agisce su recettori localizzati sulle cellule endoteliali; la sua azione provoca la liberazione di una piccola molecola che diffonde nello strato muscolare adiacente e ne causa il rilassamento.

Questa misteriosa molecola, ben presto chiamata fattore di rilassamento derivato dall'endotelio o EDRF, era quasi

I neuroni colorati in bruno nella microfotografia contengono ossido di azoto sintetasi (NOS), l'enzima che trasforma l'arginina in ossido di azoto. I neuroni colorati in grigio sono privi di NOS.

L'interferone gamma e i lipopolisaccaridi, molecole stimolatrici del sistema immunitario, trasmettono segnali al nucleo di un macrofago inducendolo a produrre ossido di azoto sintetasi. L'ossido di azoto (NO) così liberato distrugge le cellule neoplastiche inibendo sia il ciclo di Krebs, che fornisce energia alla cellula, sia il trasporto degli elettroni e la sintesi del DNA.

impossibile da identificare. Molti ricercatori, fra cui Furchgott e Louis J. Ignarro dell'Università della California a Los Angeles, tentarono invano di isolare questo composto instabile; fecero però ugualmente una scoperta significativa dimostrando che l'EDRF stimola la formazione di guanosinmonofosfato (GMP) ciclico, un «secondo messaggero» per neurotrasmettitori e ormoni. Il GMP ciclico è correlato a un secondo messaggero più noto, l'adenosinmonofosfato (AMP) ciclico.

Nel frattempo veniva perseguito un altro filone di ricerca, che avrebbe avuto importanza tutt'altro che secondaria sul lavoro di Furchgott e Ignarro. Alcuni studiosi, fra i quali Ferid Murad degli Abbott Laboratories, cercavano di verificare l'efficacia della nitroglicerina (o trinitrina) come trattamento per l'infarto miocardico; è noto che il farmaco dilata le arterie coronarie e le vene. La nitroglicerina, il componente attivo della dinamite scoperto da Alfred Nobel, era già nota alla fine dell'Ottocento per le sue proprietà terapeutiche. Lo stesso Nobel, che soffriva di angina pectoris, scrisse a un amico: «Sembra quasi un'ironia della sorte che il mio medico debba prescrivermi nitroglicerina per via orale». L'efficacia terapeutica di questo farmaco fece sì che ne venissero prodotti numerosi derivati (nitroderivati) che rimangono tuttora le sostanze più utilizzate nel trattamento dell'angina pectoris.

Sebbene Nobel avesse scoperto la nitroglicerina quasi un secolo prima, solo verso la fine degli anni settanta si cominciarono a comprendere i meccanismi molecolari della sua azione. Murad, allora alla Stanford University, scoprì che la nitroglicerina e i nitroderivati, di per sé inattivi, inducono il rilassamento dei vasi sanguigni quando sono trasformati metabolicamente in ossido di azoto. Inoltre, questo gas rilassa i muscoli stimolando la formazione di GMP ciclico, proprio come fa l'EDRF. Le due linee di ricerca, quella sull'EDRF e quella sulla nitroglicerina, sembravano dunque essersi riunite.

Nel 1986 sia Furchgott sia Ignarro erano ormai giunti alla conclusione che l'ossido di azoto o qualche suo derivato strettamente correlato potessero spiegare l'attività dell'EDRF; e finalmente, nel 1987, fu fornita la prova dell'identità delle due molecole. Salvador Moncada e collaboratori dei Wellcome Research Laboratories di Beckenham (Inghilterra) stimolarono la liberazione di EDRF dalle cellule endoteliali e ne controllarono l'effetto rilassante sulla muscolatura liscia. Al tempo stesso misurarono la quantità di ossido di azoto liberato dall'endotelio e rilevarono che essa era abbastanza elevata da giustificare completamente il rilassamento delle cellule muscolari adiacenti; si doveva dunque concludere che l'ossido di azoto fosse l'EDRF. Il gruppo di Ignarro ottenne ben presto risultati analoghi. Oltre a provocare il rilassamento dei vasi sanguigni, l'ossido di azoto inibisce la coagulazione del sangue impedendo l'aggregazione delle piastrine ed è anche il normale regolatore dell'erezione del pene.

Oggi si sa che il ruolo dell'ossido di azoto nel sistema vascolare è ancora più ampio. Sebbene si sia supposto che altre sostanze, come l'angiotensina e la noradrenalina, siano i principali determinanti della pressione sanguigna, l'ossido di azoto ne è, a quanto pare, il regolatore più importante. Diversi studiosi hanno somministrato inibitori dell'enzima che produce ossido di azoto - l'ossido di azoto sintetasi - sia ad animali sia a soggetti umani, osservando che questo trattamento provoca un aumento della pressione sanguigna rapido e più rilevante della modificazione causata da farmaci che interferiscono con la noradrenalina e l'angiotensina. Alterazioni nella regolazione dell'ossido di azoto potrebbero essere associate all'ipertensione o ad altre anomalie della pressione sanguigna.

Il susseguirsi delle scoperte sull'ossido di azoto ha ormai assunto una cadenza vertiginosa. L'identificazione del suo ruolo nei sistemi immunitario e vascolare è di per sé relativamente nuova, ma vi è una notizia ancora più recente: l'individuazione della funzione cerebrale dell'ossido di azoto. Il primo indizio che questo composto fosse attivo nel sistema nervoso giunse solo nel 1982, quando Takeo Deguchi del Tokyo Metropolitan Institute for Neurosciences notò che la formazione di GMP ciclico nel cervello richiede arginina. Naturalmente a quell'epoca nessuno sapeva che l'ossido di azoto era un messaggero chimico o che veniva sintetizzato a partire dall'arginina. Nel 1989 Moncada ipotizzò che il ruolo dell'arginina nella formazione del GMP ciclico nel cervello deve essere correlato alla produzione di ossido di azoto; e a riprova di questo osservò attività di sintesi dell'ossido di azoto in preparazioni di tessuto cerebrale.

Nello stesso tempo, si andava accumulando un'altra serie di dati. John Garthwaite dell'Università di Liverpool notò che stimolando il tessuto cerebrale con la somministrazione di glutammato si formava una sostanza a vita breve avente le stesse proprietà dell'ossido di azoto. Il glutammato, un amminoacido che funge da neurotrasmettitore eccitatorio, assicura la trasmissione sinaptica in un numero di siti cerebrali maggiore di quello di ogni altro neurotrasmettitore. I suoi effetti sono mediati da diversi sottotipi di recettori, il meglio caratterizzato dei quali è il recettore NMDA, abbreviazione di N-metil-D-aspartato, un amminoacido che agisce selettivamente in questo sottotipo di recettore del glutammato. In corrispondenza dei recettori NMDA, il glutammato apre i canali del calcio che regolano la trasmissione dei segnali nervosi, inviando così un intenso impulso eccitatorio.

In presenza di grandi quantità di glutammato, tuttavia, l'apertura di questi canali può provocare danni. Per esempio la distruzione dei neuroni durante un ictus cerebrale potrebbe essere il risultato di una «cascata del glutammato», una serie di reazioni che agiscono sulle cellule già private di ossigeno. La sollecitazione legata alla maggiore frequenza di scarica, causata dalla stimolazione a opera del glutammato, finisce per uccidere le cellule (si veda l'articolo *La terapia dell'ictus cerebrale* di Justin A. Zivin e Dennis W. Choi in «Le Scienze» n. 277, settembre 1991). Il danno ai neuroni in animali in cui sia stato indotto sperimentalmente un ictus può essere in buona parte rimediato con farmaci che blocchino i recettori NMDA. L'industria farmaceutica ha reagito tempestivamente a questa scoperta e molte ditte stanno mettendo a punto sostanze di questo tipo per il trattamento dell'ictus cerebrale.

L'osservazione fatta da Garthwaite che stimolando i recettori NMDA si provoca la liberazione di ossido di azoto indicava che questo agente fosse un mediatore del glutammato. Perciò, all'inizio del 1989, decidemmo di studiare la possibilità che l'ossido di azoto avesse un ruolo nella funzione sinaptica. Poiché si sapeva già che questo composto agisce nei vasi sanguigni tramite il GMP ciclico, il nostro obiettivo era una regione cerebrale nella quale il glutammato influisse sicuramente sul GMP ciclico. Per fortuna, negli anni settanta, James Ferrendelli della Washington University aveva scoperto che l'aggiunta di glutammato a sezioni di tessuto del cervelletto provocava un rapido e rilevante incremento del GMP ciclico.

Utilizzando analoghe preparazioni ottenute dal cervelletto, abbiamo messo a punto una tecnica per misurare l'attività dell'enzima che produce ossido di azoto, l'ossido di azoto sintetasi. Poiché l'arginina si trasforma in proporzioni eguali in ossido di azoto e citrullina, abbiamo seguito la conversione in citrullina di arginina marcata radioattivamente. La quantità misurata di citrullina doveva corrispondere esattamente a quella di ossido di azoto. Con questo metodo abbiamo trovato che l'attività enzimatica risultava triplicata quando si aggiungeva NMDA o glutammato. Fummo impressionati dall'estrema rapidità di questo effetto, che avveniva in pochi secondi. La scoperta era abbastanza sorprendente perché per aumentare l'attività di un enzima occorre di solito parecchio tempo.

Lavorando sulle stesse sezioni di tessuto, confermammo che l'NMDA provoca un notevole aumento dei livelli di GMP ciclico e decidemmo di verificare se vi fosse un nesso causale fra la formazione di ossido di azoto e quella di GMP ciclico. Trovammo facilmente la risposta aggiungendo alla preparazione metilarginina - l'inibitore dell'ossido di azoto sintetasi - e osservando che essa blocca la formazione del GMP ciclico se presente nelle stesse concentrazioni alle quali inibisce l'ossido di azoto sintetasi. Dato che entrambi i processi sono bloccati dalla stessa quantità di metilarginina, si può concludere che sono strettamente correlati. Anche Garthwaite e Moncada hanno osservato il blocco della formazione del GMP ciclico.

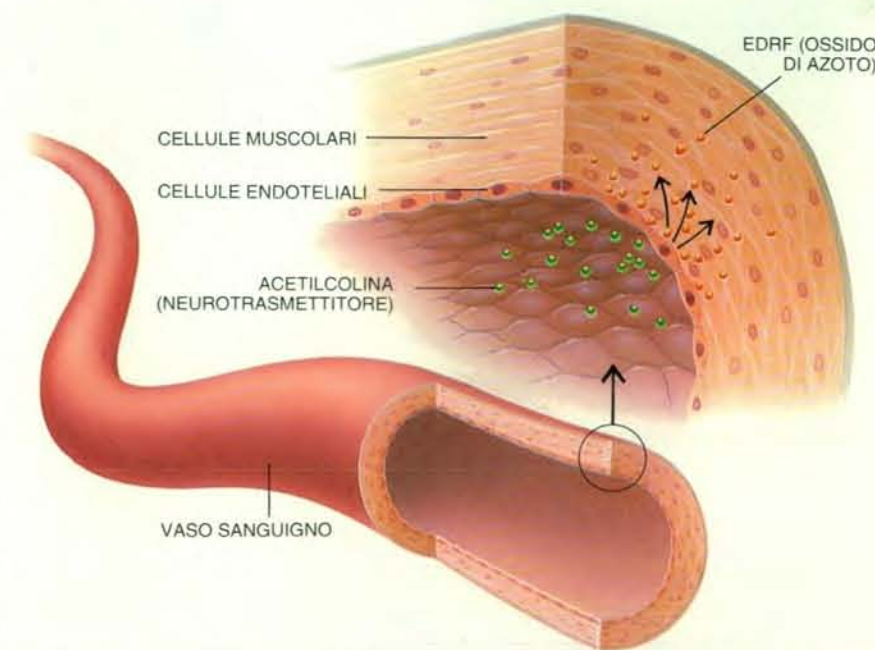
Il passo successivo fu quello di stabilire in quali regioni cerebrali sia attivo l'ossido di azoto. Normalmente i neurologi ricavano indizi sulle funzioni cerebrali di una molecola determinandone gli specifici siti di azione neuronici, a loro volta associati a particolari vie e funzioni biologiche. Ma cercare di localizzare una molecola a vita breve come l'ossido di azoto sembrava vano; tentammo allora di individuare l'ossido di azoto sintetasi.

Uno dei metodi più efficienti per localizzare le proteine è fornito dall'immunoistochimica. Ai campioni di tessuto si applica un anticorpo diretto contro la molecola da esaminare e questo si lega al proprio antigene, ossia alla molecola in questione. Varie tecniche, fra cui la colorazione selettiva, vengono poi utilizzate per marcare i siti dove l'anticorpo si è legato all'antigene.

Prima di poter procedere con i nostri esperimenti immunoistochimici, tuttavia, dovevamo ottenere anticorpi contro l'ossido di azoto sintetasi; e prima ancora, ovviamente, dovevamo isolare l'enzima stesso. Compito tutt'altro che facile: provammo numerose tecniche di purificazione, ma in ogni caso l'attività enzimatica veniva rapidamente annullata. Ci rendemmo allora conto che un qualche aspetto del processo di purificazione sottraeva una sostanza chimica che fungeva da cofattore per l'attività dell'enzima. Perciò considerammo la possibilità di purificare questo ipotetico cofattore.

Dato che questo approccio appariva piuttosto complicato, cercammo di escogitare qualche idea più promettente. Moncada aveva scoperto che la sintesi dell'ossido di azoto richiede la presenza di calcio. Questo ione agisce sovente legandosi a un cofattore comunissimo, la calmodulina, e, in effetti, aggiungendo una piccola quantità di calmodulina ad alcune delle nostre preparazioni, notammo subito un netto incremento dell'attività enzimatica. Con un autentico colpo di fortuna, avevamo trovato il misterioso cofattore dell'ossido di azoto sintetasi.

Una volta riconosciuta l'associazione fra ossido di azoto, calcio e calmodulina, potemmo procedere alla purificazione dell'enzima. Cosa ancora più importante, eravamo ora in grado di spiegare perché i recettori NMDA avviassero così rapidamente la sintesi dell'ossido di azoto dopo essere stati stimolati dal glutammato. Come si sa, il glutammato attiva la trasmissione sinaptica in corrispondenza dei recettori NMDA aprendo i canali ionici che promuovono il passaggio di ioni calcio dall'esterno all'interno dei neuroni. Pertanto il glutammato fa sì che il calcio entri nelle cellule; legandosi alla calmodulina, gli ioni attivano l'ossido di azoto sintetasi. L'intero processo richiede non più di qualche millisecondo.

Purificata l'ossido di azoto sintetasi, iniziammo a mettere a punto anticorpi

Un vaso sanguigno si dilata quando un neurotrasmettitore come l'acetilcolina si lega alle cellule endoteliali. Queste cellule liberano il fattore di rilassamento derivato dall'endotelio (EDRF), che diffonde nelle cellule muscolari adiacenti e ne provoca il rilassamento. Nel 1987 si è scoperto che l'EDRF è in realtà l'ossido di azoto.

contro di essa insieme al laureando Paul M. Hwang. Verificammo così la presenza dell'enzima nel cervello e nel resto dell'organismo. La nostra osservazione più sorprendente fu che l'ossido di azoto sintetasi si trova quasi esclusivamente nei neuroni. Questi costituiscono solo il 15 per cento circa di tutte le cellule cerebrali; il restante 85 per cento è dato dalle cellule gliali che provvedono alle necessità metaboliche e di sostegno dei neuroni. Tuttavia nelle cellule gliali non sono state rilevate tracce dell'enzima.

È interessante il fatto che l'ossido di azoto sintetasi si trovi solo in determinate popolazioni di neuroni. Nell'ipofisi, per esempio, l'enzima è presente nei neuroni il cui corpo cellulare è contenuto nell'ipotalamo, a partire dal quale essi si estendono nel lobo posteriore dell'ipofisi. Questi neuroni hanno la specifica funzione di sintetizzare due ormoni, la vasopressina e l'ossitocina. Nella ghiandola surrenale l'ossido di azoto sintetasi è fortemente concentrata in una rete di neuroni che stimolano le cellule surrenali a secernere adrenalina. Nell'intestino l'enzima è contenuto in un raggruppamento di neuroni che prende il nome di plesso mioenterico e ha una funzione di regolazione della peristalsi. Nella corteccia cerebrale l'ossido di azoto sintetasi si trova solo nel due per cento circa dei neuroni. Al di fuori del sistema nervoso l'enzima è presente nello strato endoteliale che riveste i vasi sanguigni.

La situazione da noi osservata è straordinaria proprio per il fatto che l'ossido di azoto sintetasi è contenuta soprattutto nei neuroni. Le funzioni dell'ossido di azoto nelle cellule nervose devono perciò avere importanza paragonabile, se non superiore, a quelle svolte nei macrofagi e nei vasi sanguigni.

Sebbene fossimo riusciti a individuare i siti cerebrali in cui viene prodotto l'ossido di azoto, le nostre scoperte non indicavano con chiarezza la funzione di questo composto, la cui localizzazione appariva misteriosa in quanto non corrispondeva a quella di alcun neurotrasmettitore noto. Ben presto però notammo un fatto importante: la peculiare localizzazione dell'ossido di azoto sintetasi nella corteccia cerebrale corrisponde a una specifica disposizione di neuroni individuabili con un certo colorante.

Questa tecnica di colorazione fu messa a punto a metà degli anni sessanta dall'istochimico Anthony Pearse dell'Università di Londra. Trattando sezioni cerebrali con blu di tetrazolio, Pearse osservò che certi neuroni assumevano una viva colorazione blu per aggiunta di un coenzima, il nicotinammideadenindinucleotidefosfato ridotto (NADPH). Questi neuroni, detti diaforasici, costituiscono circa il 2 per cento della corteccia cerebrale. Il NADPH funge da donatore di elettroni per gli enzimi ossidanti, e quindi si supponeva che la diaforasi mediasse un qualche tipo di ossidazione.

In un primo momento, però, ben pochi ricercatori si interessarono a questa tecnica. Verso la metà degli anni ottanta M. Flint Beal e Joseph B. Martin del Massachusetts General Hospital osservarono che i neuroni specificamente colorati dal blu di tetrazolio erano selettivamente resistenti agli effetti degenerativi di certe malattie. Nella corea di Huntington si ha la degenerazione anche del 95 per cento dei neuroni del nucleo caudato, ma i neuroni diaforasici rimangono pressoché intatti. Anche nell'infarto cerebrale e in alcune aree cerebrali colpite dalla malattia di Alzheimer i neuroni diaforasici appaiono ugualmente resistenti. L'effetto neurotossico esercitato dall'NMDA su neuroni in coltura, che viene utilizzato come modello per l'ictus, può distruggere anche il 90 per cento delle cellule, ma i neuroni diaforasici non vengono minimamente danneggiati.

Il motivo per cui i neuroni diaforasici sono in grado di resistere a questi attacchi neurotossici incuriosiva molti ricercatori, anche perché la soluzione di questo mistero avrebbe grande importanza per il trattamento di gravissime patologie cerebrali, fra cui l'ictus, la malattia di Alzheimer e la corea di Huntington.

La sovrapposizione fra neuroni diaforasici e neuroni cerebrali contenenti ossido di azoto sintetasi era particolarmente affascinante. Data l'importanza del possibile rapporto fra ossido di azoto e NADPH, ci mettemmo immediatamente all'opera insieme a Hwang, a Ted M. Dawson e a Majid Fotuhi e ben presto riuscimmo a dimostrare che, sia nel cervello sia nei tessuti periferici, esattamente gli stessi neuroni venivano evidenziati dalla colorazione per l'ossido di azoto sintetasi e per la diaforasi. Spiegammo anche quale fosse il rapporto fra l'attività dell'ossido di azoto sintetasi e il colorante specifico per la diaforasi: quest'ultimo acquisisce gli elettroni che normalmente vengono usati per ossidare l'arginina a ossido di azoto, dando così origine a una colorazione blu.

La nostra scoperta, per quanto entusiasmante, sembrava illogica. È chiaro che un qualche effetto dovuto alla sintesi dell'ossido di azoto permette ai neuroni di resistere a danni neurotossici, ma nello stesso tempo l'ossido di azoto è un prodotto dell'attività del glutammato e quest'ultimo è responsabile della neurotossicità. Se si conoscesse il meccanismo molecolare a cui è dovuta la resistenza, si potrebbe risolvere il paradosso e comprendere meglio perché i neuroni che mancano dell'ossido di azoto sintetasi soccombano così facilmente agli effetti neurotossici.

Un modo per aggirare il dilemma è di proporre che i neuroni che sintetizzano ossido di azoto lo liberino e che questo, a causa della sua tossicità, distrugga i neuroni adiacenti. Sapevamo già che, dopo una stimolazione con livelli moderati di glutammato, l'ossido di azoto provoca la formazione di GMP ciclico nelle cellule nervose adiacenti. Ma forse, in presenza di alti livelli di glutammato, i neuroni che producono ossido di azoto si comportano come i macrofagi, vale a dire ne liberano in quantità tossiche. Se questa teoria è corretta, allora ci si può attendere che gli inibitori dell'ossido di azoto sintetasi prevengano la neurotossicità.

Per affrontare direttamente il problema, Valina L. Dawson ed Edythe D. London dell'Addiction Research Center di Baltimora, insieme a Dawson e a noi, hanno esaminato la neurotossicità dell'NMDA in colture di neuroni della corteccia cerebrale. In questo sistema modello, messo a punto da Dennis W. Choi, allora alla Stanford University, si aggiunge NMDA a colture ottenute da cellule cerebrali di feto di ratto. Il giorno successivo a un'esposizione di soli cinque minuti all'NMDA anche il 90 per cento dei neuroni appare distrutto. Questo modello rispecchia la neurotossicità che si ha nell'infarto cerebrale.

Nei nostri esperimenti su queste colture abbiamo osservato che la nitroarginina, un inibitore particolarmente potente e selettivo dell'ossido di azoto sintetasi, previene completamente la neurotossicità indotta dall'NMDA. Anche l'eliminazione dell'arginina dal mezzo di incubazione ha un effetto protettivo sulle cellule. Un altro inibitore della tossicità è l'emoglobina, che si lega all'ossido di azoto e lo inattiva.

È quindi chiaro che in coltura l'ossido di azoto è responsabile della neurotossicità prodotta dal glutammato in quanto agisce a livello dei recettori NMDA. Dato che gli antagonisti dell'NMDA possono bloccare il danno indotto dal glutammato che è associato all'infarto cerebrale, può darsi che l'ossido di azoto moduli anche la distruzione dei neuroni che si ha in questa patologia.

Bernard Scatton e colleghi della Synthelabo di Parigi hanno recentemente inoculato piccole dosi di nitroarginina in alcuni topi subito dopo aver indotto in essi un ictus. La nitroarginina ha limitato del 73 per cento il danno alle cellule; in un esperimento parallelo, tuttavia, un potente antagonista dell'NMDA ha avuto un effetto protettivo solo del 55 per cento. Queste osservazioni indicano che gli inibitori dell'ossido di azoto sintetasi potrebbero avere utili effetti terapeutici in caso di ictus e di danno neurologico associato a un eccesso di glutammato.

La funzione esatta dell'ossido di azoto nei neuroni deve essere ancora chiarita, e sebbene si siano fatti dei progressi, il mistero continua ad approfondirsi. Diverse serie di dati indicano che l'ossido di azoto ha un'azione molto simile a quella di neurotrasmettitori ben noti. Nell'intestino, per esempio, il rilassamento muscolare correlato alla peristalsi è mediato dal plesso mioenterico, ma il neurotrasmettitore responsabile non è ancora stato identificato. L'aver dimostrato che l'ossido di azoto sintetasi si concentra in questi neuroni fa pensare che il misterioso messaggero possa essere l'ossido di azoto. In diversi laboratori è stato poi dimostrato che il normale rilassamento muscolare prodotto dalla stimolazione elettrica dei neuroni del plesso mioenterico può essere bloccato dagli inibitori dell'ossido di azoto sintetasi. Inoltre l'introduzione diretta di ossido di azoto nell'intestino simula gli effetti della stimolazione nervosa.

Se l'ossido di azoto è un neurotrasmettitore, il suo comportamento è davvero inconsueto. I neurotrasmettitori sono di solito sostanze stabili immagazzinate nelle vescicole sinaptiche delle terminazioni nervose. Liberati nel corso di un processo in cui la vescicola sinaptica si fonde con la membrana del neurone, i neurotrasmettitori interagiscono con recettori proteici posti sulla superficie della membrana dei neuroni adiacenti.

Invece l'ossido di azoto non è immagazzinato nelle vescicole. In collaborazione con Marcello Costa della Flinders University in Australia, abbiamo utilizzato la microscopia elettronica per vi-

L'ossido di azoto è liberato nei neuroni quando il glutammato si lega al recettore NMDA (*in alto*). Il legame fa sì che gli ioni calcio entrino nel neurone, dove si legano alla calmodulina che attiva la NOS. Questa trasforma l'arginina in citrullina e NO.

L'attivazione del recettore fa aumentare il GMP ciclico - un secondo messaggero - e la citrullina (*in basso a sinistra*). Quando il recettore è bloccato dalla metilarginina (*in basso a destra*), il livello sia di questi composti sia del NO diminuisce.

L'ossido di azoto sintetasi in questo cervello di ratto si evidenzia nelle zone chiare del bulbo olfattivo, del corpo striato (l'area più colpita dalla corea di Huntington), dei collicoli (dove sono elaborate le informazioni visive e uditive) e del cervelletto.

OSSIDO DI AZOTO SINTETASI

CITOCROMO P-450 RIDUTTASI

FLAVINMONONUCLEOTIDE

FLAVINADENIN-DINUCLEOTIDEPIROFOSFATO

FLAVINADENIN-DINUCLEOTIDEISOALLOSSAZINA

NICOTINAMMIDEADENINDINUCLEOTIDE-FOSFATORIBOSIO RIDOTTO

NICOTINAMMIDEADENINDINUCLEOTIDE-FOSFATOADENINA RIDOTTA

Le sequenze di amminoacidi ricavate da segmenti di DNA della NOS e della citocromo P-450 riduttasi fanno pensare che i due enzimi abbiano analoga attività. Entrambi possiedono siti di legame per certi coenzimi (*cornici in rosso*). I trattini verticali indicano amminoacidi identici nei due enzimi; i pallini corrispondono ad amminoacidi molto simili.

sualizzare l'ossido di azoto sintetasi nei neuroni del plesso mioenterico e abbiamo scoperto che l'enzima è localizzato nel citoplasma, non nelle vescicole sinaptiche. È quindi probabile che l'ossido di azoto venga sintetizzato «su richiesta» nel neurone e venga liberato per semplice diffusione dalla terminazione nervosa. Invece di agire a livello di un recettore di membrana, l'ossido di azoto diffonde in un neurone adiacente: il suo bersaglio è l'atomo di ferro presente nel sito attivo dell'enzima che forma il GMP ciclico. Legandosi al ferro, l'ossido di azoto fa variare la configurazione tridimensionale dell'enzima, che aumenta la propria attività e la produzione di GMP ciclico. Dato il singolare meccanismo di azione, l'ossido di azoto rappresenta una classe completamente nuova di messaggero chimico tra neuroni.

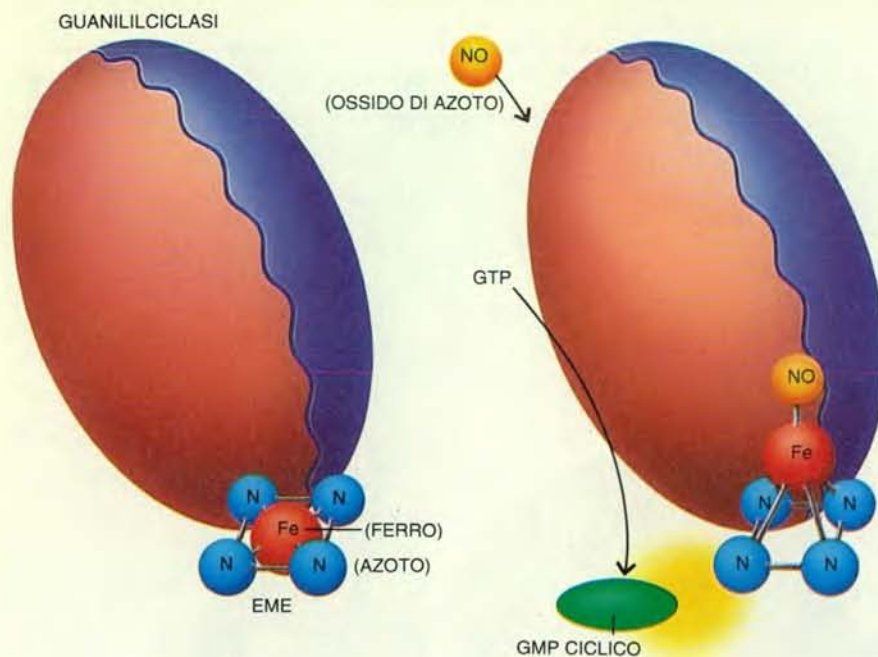
L'ossido di azoto potrebbe avere un ruolo anche nei processi che stanno alla base dell'apprendimento e della memoria. La maggior parte dei ricercatori è portata a ritenere che la memoria consista in un aumento o in una diminuzione a lungo termine della trasmissione attraverso certe sinapsi dopo che i neuroni sono stati ripetutamente stimolati (si veda l'articolo *Il significato dei sogni* di Jonathan Winson in «Le Scienze» n. 269, gennaio 1991). Pertanto, in esperienze sulla memoria come il potenziamento (LTP) o la depressione della trasmissione sinaptica (entrambi a lungo termine), gli sperimentatori stimolano più volte particolari neuroni e poi rilevano incrementi o diminuzioni persistenti nella trasmissione tra le sinapsi.

Qual è il ruolo dell'ossido di azoto sintetasi in questi processi a lungo termine? Eric R. Kandel e collaboratori della Columbia University, Daniel Madison e colleghi della Stanford University e Georg Böhme e collaboratori della Rhône-Poulenc Rorer in Francia hanno studiato indipendentemente l'azione degli inibitori dell'ossido di azoto sintetasi determinando gli effetti di queste sostanze sull'LTP nell'ippocampo, una regione cerebrale che partecipa all'elaborazione della memoria; in tutti gli esperimenti si è osservato che gli inibitori bloccavano effettivamente l'LTP. Inoltre Katsuei Shibuki e Daisuke Okada del Laboratory for Neural Networks in Giappone hanno dimostrato che questa inibizione avviene anche nel cervelletto.

Ci siamo chiesti se l'ossido di azoto potesse essere solo il primo membro di una famiglia del tutto nuova di messaggeri chimici fra neuroni. Alcune conferme di questa possibilità sono state fornite dalla clonazione dell'ossido di azoto sintetasi, che abbiamo effettuato insieme a Hwang, Charles E. Glatt e Charles Lowenstein del nostro laboratorio e in collaborazione con Randall R. Reed dell'Howard Hughes Medical Institute della Johns Hopkins University.

Dopo aver stabilito la sequenza degli amminoacidi, l'abbiamo confrontata con quella di tutte le proteine note e abbiamo riscontrato somiglianze con un solo enzima conosciuto dei mammiferi: la citocromo P-450 riduttasi. Questa scoperta era sorprendente in quanto la citocromo P-450 riduttasi non era mai stata associata all'encefalo o alla funzione dei neurotrasmettitori.

Questo enzima funge invece da donatore di elettroni per gli enzimi P-450 del fegato, i quali metabolizzano un'ampia varietà di farmaci e, in alcuni organismi, sono attivati da inquinanti come la diossina e certi idrocarburi aromatici. Le funzioni di donatore di elettroni della citocromo P-450 riduttasi sono mediate da tre coenzimi, e quindi la sequenza amminoacidica dell'enzima presenta siti di riconoscimento per ciascuno di questi: NADPH, flavinadenindinucleotide e flavinmononucleotide.

La cosa stupefacente è che anche l'ossido di azoto sintetasi possiede siti di riconoscimento per gli stessi tre coenzimi, e per giunta collocati nei medesimi punti

Il recettore dell'ossido di azoto è l'atomo di ferro della guanililciclasi (*a sinistra*), un enzima che favorisce la sintesi di GMP ciclico. Il legame con ossido di azoto fa sì che il gruppo eme contenente ferro alteri la propria configurazione tridimensionale (*a destra*); questa alterazione promuove la conversione di GTP in GMP ciclico.

in cui si trovano nella citocromo P-450 riduttasi. Vi è anche una considerevole somiglianza nelle sequenze amminoacidiche globali dei due enzimi.

L'ossido di azoto sintetasi, tuttavia, ha diversi altri siti di legame che mancano nella citocromo P-450 riduttasi: per esempio possiede un sito di legame per la calmodulina e uno dove avviene la fosforilazione. (Quest'ultimo processo consiste nell'aggiunta di gruppi fosfato alle proteine per azione di una famiglia di enzimi denominati proteinchinasi.) I lavori di Paul Greengard della Rockefeller University e di molti altri ricercatori hanno permesso di stabilire che la fosforilazione proteica trasmette informazioni provenienti dai secondi messaggeri, come l'AMP ciclico, a varie proteine intracellulari. La fosforilazione è uno dei principali meccanismi di segnalazione all'interno della cellula. Avevamo in precedenza osservato che l'ossido di azoto sintetasi può essere fosforilata da una proteinchinasi dipendente dall'AMP ciclico, da una proteinchinasi dipendente dal sistema calcio-calmodulina e da un enzima denominato proteinchinasi C. Queste chinasi sono importanti enzimi di fosforilazione che regolano la ricettività cellulare nei confronti di ormoni, neurotrasmettitori e fattori di crescita.

Dato che l'ossido di azoto sintetasi possiede tanti siti deputati alla regolazione, deve ovviamente essere influenzata da un numero di fattori più elevato rispetto alla maggior parte degli altri enzimi. Questa ampia capacità di risposta appare logica se si considera quanto sia cruciale il ruolo dell'ossido di azoto nella modulazione fine di un gran numero di processi. La sua funzione sembra in contrasto con l'attività più banale svolta dalla citocromo P-450 riduttasi, a dispetto delle molte somiglianze fra le sequenze di amminoacidi dei due enzimi.

Quali potrebbero essere allora i punti in comune? Oltre a regolare gli enzimi che metabolizzano i farmaci, la citocromo P-450 riduttasi dona elettroni all'emeossigenasi, l'enzima che scinde l'eme (il costituente dei globuli rossi che trasporta l'ossigeno). La forma meglio caratterizzata di emeossigenasi, denominata numero uno, si concentra nella milza, dove di norma si ha la degradazione dei globuli rossi, ma un'altra forma meno conosciuta dell'enzima, la numero due, è presente in elevate concentrazioni nel cervello. Dato che la presenza dell'emeossigenasi numero due coincide costantemente con quella della citocromo P-450 riduttasi, sembra probabile che una delle funzioni principali di quest'ultima sia proprio la donazione di elettroni alla forma numero due.

Un attento esame del processo di donazione di elettroni ha rivelato quello che potrebbe essere il secondo membro della nuova famiglia di messaggeri tra neuroni. Durante questo processo l'emeossigenasi libera monossido di carbonio, un gas tossico che, come tutti sanno, è un residuo di combustione; legandosi all'emoglobina, esso ne pregiudica l'attività di trasporto di ossigeno. Tuttavia il monossido di carbonio è una sostanza normalmente presente nell'organismo e, stabilendo un legame con il gruppo eme della guanililciclasi, l'enzima che induce la sintesi di GMP ciclico, stimola la produzione di questo secondo messaggero esattamente come fa l'ossido di azoto.

Ajay Verma, del nostro laboratorio, ha proposto il monossido di carbonio come eccellente candidato al ruolo di messaggero fra neuroni e con David Hirsch ha raccolto numerosi indizi a sostegno dell'ipotesi. Verma e Hirsch hanno dimostrato che l'RNA messaggero per l'emeossigenasi numero due si trova in popolazioni di neuroni la cui distribuzione segue da vicino quella della citocromo P-450 riduttasi e della guanililciclasi.

Queste scoperte sono coerenti con la possibilità che il monossido di carbonio sia il principale regolatore del GMP ciclico nel cervello. Trattando colture di neuroni cerebrali con un inibitore dell'emeossigenasi, Verma e Hirsch hanno osservato una radicale diminuzione dei livelli di GMP ciclico; viceversa gli inibitori dell'ossido di azoto sintetasi non hanno avuto alcun effetto. Pertanto si può desumere che il monossido di carbonio mantenga il GMP ciclico a determinati livelli in varie regioni cerebrali. Le somiglianze strutturali dell'ossido di azoto sintetasi e della citocromo P-450 riduttasi sembrerebbero rispecchiare un ruolo analogo nella sintesi dei due messaggeri «fratelli», il monossido di carbonio e l'ossido di azoto. Sebbene questi due composti si trovino in regioni diverse del cervello e partecipino a funzioni biochimiche differenti, per meccanismo di azione sono molto simili.

Che il monossido di carbonio faccia parte o meno di una nuova classe di neurotrasmettitori, è comunque evidente che l'ossido di azoto è un messaggero chimico importante e straordinario. Piccolo, di breve vita e dallo strano comportamento, sembra poter rivaleggiare per il suo significato biochimico con i principali neurotrasmettitori.

BIBLIOGRAFIA

MONCADA S. e HIGGS E. S. (a cura), *Nitric Oxide from L-Arginine: A Bioregulatory System*, Excerpta Medica, 1990.

BREDT DAVID S., HWANG PAUL M., GLATT CHARLES E., LOWENSTEIN CHARLES, REED RANDALL R. e SNYDER SOLOMON H., *Cloned and Expressed Nitric Oxide Synthase Structurally Resembles Cytochrome P-450 Reductase* in «Nature», 351, n. 6329, 27 giugno 1991.

MONCADA S., PALMER R. M. J. e HIGGS E. A., *Nitric Oxide: Physiology, Pathophysiology and Pharmacology* in «Pharmacological Reviews», 43, n. 2, giugno 1991.

BREDT D. S. e SNYDER S. H., *Nitric Oxide: A Novel Neuronal Messenger* in «Neuron», 8, n. 1-2, gennaio 1992.

Le nebulose planetarie

Sono nubi fluorescenti di gas che, rappresentando l'ultimo sussulto di vita degli astri simili al Sole, sono di aiuto agli astronomi per capire la dinamica evolutiva delle stelle e forse anche il destino dell'universo

di Noam Soker

Come le persone, anche le stelle invecchiano e muoiono: quelle di massa maggiore hanno una vita breve e scompaiono in una pirotecnica esplosione di supernova, ma le stelle di massa intermedia come il Sole, molto più comuni, concludono i propri giorni in maniera ben più elegante. Quando esauriscono il proprio combustibile nucleare, si gonfiano e, in un ultimo sussulto, espellono i propri strati più esterni creando una nube sferica di gas in espansione. Il nucleo piccolo ma rovente, che è quel che rimane della stella, scalda il gas fino a farlo brillare. Per alcune migliaia di anni, quindi, la stella si circonda di una splendida nube scintillante: una nebulosa planetaria.

Le nebulose planetarie si presentano in una gamma spettacolare di forme, dimensioni e strutture, che racchiude indizi affascinanti sulla natura dell'evoluzione stellare e sulla miriade di forme che può assumere l'interazione fra le stelle e il loro ambiente. Per esempio parte degli elementi, come carbonio, azoto e ossigeno, che si formano nel nucleo delle stelle, migra verso gli strati più esterni. L'abbondanza relativa di questi elementi in una nebulosa planetaria può quindi rivelare in che misura la materia delle varie parti di una stella si mescoli. La nebulosa in espansione si fonde col gas interstellare e lo arricchisce di questi elementi pesanti: è il processo principale che restituisce allo spazio interstellare la materia concentratasi nelle stelle.

Da qualche anno gli astronomi utilizzano le nebulose planetarie, che sono brillanti e relativamente facili da identificare, per perfezionare le misurazioni della distanza di galassie lontane. Questi calcoli aiutano i cosmologi a determinare l'età e il destino ultimo dell'universo.

Le nebulose planetarie non si vedono a occhio nudo. Il loro studio, anzi, è stato perseguito scrupolosamente solo molto tempo dopo l'invenzione del telescopio. L'osservazione delle nebulose planetarie viene ostacolata dal fatto che questi oggetti hanno una luminosità superficiale molto bassa e quindi tendono a confondersi con lo sfondo del cielo. La luce della nebulosa Elica (NGC 7293), per esempio, si disperde su un'area pari a quella della Luna piena.

Negli ultimi 20 anni del XVIII secolo il grande astronomo inglese William Herschel diede inizio a uno studio approfondito e a una classificazione delle nebulose, chiazze sfumate di luce che non hanno aspetto stellare. Fu proprio Herschel, in un articolo pubblicato nel 1785, a coniare l'espressione «nebulosa planetaria» per descrivere una classe di oggetti la cui forma tondeggiante li faceva assomigliare a versioni spettrali di pianeti. (Successivamente l'elenco di Herschel, ampliato, divenne il New General Catalogue; ancora oggi le nebulose planetarie si indicano con il numero NGC, cioè il numero d'ordine all'interno di questo catalogo di oggetti celesti non stellari.) Herschel ipotizzava erroneamente che le nebulose planetarie fossero oggetti giovani, non ancora condensatisi in stelle.

Da quell'epoca il numero delle nebulose planetarie catalogate è cresciuto da una decina a più di 1600 soltanto nella nostra galassia (il numero totale è forse 20 volte maggiore). Cosa ancora più importante, gli astronomi hanno imparato molte cose sulla loro vera natura grazie alla spettroscopia, che permette di scindere la radiazione emessa da un oggetto nelle diverse componenti per analizzarne la natura in funzione delle lunghezze d'onda individuate. Ogni elemento infatti emette e assorbe radiazione soltanto in corrispondenza di certe lunghezze d'onda caratteristiche, che nello spettro appaiono come linee sottili, le righe spettrali. La lunghezza d'onda delle righe spettrali sembra diminuire se la sorgente di radiazione si avvicina all'osservatore e aumentare se essa se ne allontana. Questo fenomeno, apparentemente semplice, noto come spostamento Doppler, ha fornito e continua a fornire numerosissime informazioni sulla struttura e sulla dinamica delle nebulose planetarie.

All'inizio del XX secolo si scoprì che le righe spettrali delle nebulose planetarie hanno una tipica forma a due picchi, che indica come questi oggetti si stiano espandendo. La radiazione emessa dal lato della nebulosa che si avvicina a noi risulta infatti spostata a una lunghezza d'onda leggermente inferiore, mentre quella del lato che si allontana assume una lunghezza d'onda leggermente più grande. È chiaro quindi che non si tratta di oggetti che si contraggono per formare nuove stelle, ma di materia espulsa da astri vecchi. Nel 1956 l'astrofisico sovietico Iosif S. Šklovskij formulò e dimostrò l'interpretazione moderna, secondo la quale le nebulose si formano dall'espulsione degli strati esterni di vecchie stelle giganti rosse, e quindi rappresentano la fine, e non l'inizio, della vita di una stella.

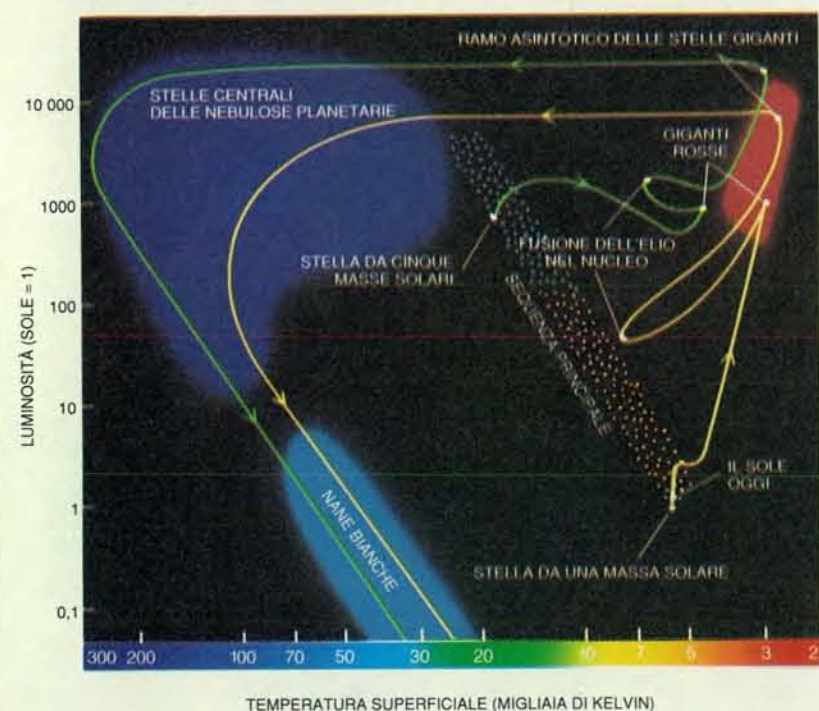
Numerose osservazioni hanno dimostrato che le nebulose planetarie si espandono a una velocità compresa tra 5 e 100 chilometri al secondo, con un valore medio di circa 20 chilometri al secondo; in alcuni casi è possibile confermare questi valori esaminando visivamente l'espansione della nebulosa. Formulando una stima delle dimensioni della nebulosa e del rallentamento della sua espansione se ne può calcolare l'età. I valori ricavati, compresi tra alcune migliaia di anni e 30 000 anni, hanno contribuito a porre vincoli ai modelli di evoluzione delle nebulose planetarie e delle stelle che ne occupano il centro.

Esaminando gli spettri alterati dallo spostamento Doppler si è scoperto inoltre che dalla stella centrale di molte nebulose planetarie spira un flusso di gas, o «vento», ad alta velocità. I venti stellari danno origine a un profilo spettrale caratteristico perché gli atomi possono assorbire radiazione alle stesse lunghezze d'onda alle quali la emettono, e quindi gli atomi interposti lungo la linea di vista assorbono in parte la luce della stella prima che questa giunga alla Terra, dando origine a una riga scura (in assorbimento) nello spettro. Dato che questi atomi si trovano in quella parte del vento che si muove verso la Terra, le righe in assorbimento risultano spostate a una lunghezza d'onda inferiore. Degli altri atomi, che non sono proiettati sull'immagine della stella centrale, si vede la radiazione in emissione e non in assorbimento. Gli atomi emittenti si muovono in tutte le direzioni e quindi presentano uno spostamento Doppler complessivo nullo.

Ne consegue che lo spettro della stella contiene una riga in emissione, ma anche una riga in assorbimento dello stesso atomo, a lunghezza d'onda leggermente inferiore. La lunghezza d'onda alla quale appare la riga in assorbimento (o più esattamente la differenza di lunghezza d'onda tra le due righe) consente di valutare la velocità del vento: si sono misurati venti molto veloci, che si muovono anche a 4000 chilometri al secondo. L'intensità delle righe in assorbimento e in emissione può essere a sua volta uti-

Le nebulose planetarie si formano quando una vecchia stella di massa simile a quella del Sole si libera dei propri strati esterni. La radiazione del residuo stellare fa brillare il gas circostante. Le nebulose sferiche, come NGC 2392 (*a*) sono l'eccezione: molto più comuni sono le forme ellittiche, come quella di NGC 7662 (*b*) e NGC 7354 (*c*). NGC 650-1 (*d*), NGC 2346 (*e*) e NGC 2440 (*f*) possiedono una forma a farfalla, mentre NGC 6543 (*g*) sfugge a qualsiasi classificazione. Intorno ad alcune nebulose si trovano addensamenti brillanti, come nel caso di NGC 7009 (*h*) e NGC 3242 (*i*). L'emissione dell'ossigeno ionizzato appare in verde, quella dell'azoto in rosso e quella dell'elio in blu. L'elemento che emette la radiazione più intensa in ciascuna regione dipende dal valore locale di temperatura, densità e quantità di radiazione stellare ricevuta.

Il ciclo di vita delle stelle di tipo solare

L'evoluzione stellare si può illustrare con un diagramma della luminosità in funzione della temperatura superficiale. Le curve indicano il percorso evolutivo di due stelle, una di massa pari a quella del Sole e l'altra di massa cinque volte maggiore. Circa il 90 per cento della vita di una stella si svolge all'interno di una zona di stabilità, la sequenza principale; le stelle di grande massa evolvono però più rapidamente di quelle meno pesanti. Quando si esaurisce l'idrogeno, entrambe le stelle si espandono e si raffreddano, trasformandosi in giganti rosse. All'inizio della fusione dell'elio le stelle si contraggono per breve tempo, ma quando l'elio nel nucleo si esaurisce, si espandono ancora e diventano giganti del ramo asintotico. In queste condizioni le stelle divengono instabili e perdono completamente i propri strati esterni nel giro di 100 000 anni. Quando il resto del nucleo si contrae ancora riscaldandosi sempre più, la radiazione emessa fa risplendere il gas circostante, dando origine a una nebulosa planetaria. Infine, esaurito il combustibile nucleare, le stelle cominciano a raffreddarsi e dopo un periodo compreso tra 10 000 e 30 000 anni non riescono più a far brillare la nebulosa, che continua a espandersi. Ormai nude, trascorrono ciò che rimane dei propri giorni in forma di resti estremamente densi, le cosiddette nane bianche.

lizzata per dedurre la densità del vento e quindi la velocità alla quale la stella perde massa, ma questo metodo è ancora poco preciso.

Le osservazioni spettroscopiche hanno rivelato anche la composizione delle nebulose planetarie. Ciascun atomo produce infatti una configurazione caratteristica di righe spettrali che si può osservare e confrontare con gli spettri misurati in laboratorio. All'inizio, però, le emissioni delle nebulose planetarie parvero misteriose: nel 1866 l'astronomo britannico William Huggins scoprì che molte nebulose planetarie mostrano tre righe spettrali di intensità particolarmente elevata, due delle quali non corrispondevano ad alcuna emissione allora conosciuta. Huggins attribuì queste righe a un nuovo elemento ipotetico che chiamò - e non poteva fare diversamente - nebulio. Nel 1927 però Ira S. Bowen del California Institute of Technology capì che le misteriose righe spettrali erano dovute in realtà a ioni (cioè atomi privi di alcuni dei propri elettroni), e identificò lo ione responsabile dell'emissione nell'ossigeno ionizzato due volte (privo cioè di due elettroni), che è estremamente instabile sulla Terra, ma è in grado di sopravvivere a lungo nel vuoto quasi totale dello spazio.

La lunghezza d'onda delle righe in emissione identifica in maniera univoca sia l'elemento sia il numero di elettroni mancanti, cioè lo stato di ionizzazione. Quasi tutta la radiazione proveniente dalle nebulose planetarie è emessa da ioni che vengono eccitati da collisioni o dalla radiazione emessa dalla stella calda, oppure che ricatturano elettroni liberi. Lawrence H. Aller dell'Università della California a Los Angeles, James B. Kaler dell'Università dell'Illinois e molti altri hanno esaminato le righe in emissione per determinare l'abbondanza relativa delle varie specie atomiche nelle nebulose planetarie, usando sia telescopi ottici sia il satellite *International Ultraviolet Explorer*, in funzione dal 1978.

La massa delle nebulose planetarie è costituita per circa il 70 per cento da idrogeno e per il 28 per cento da elio. I componenti principali del due per cento rimanente sono carbonio, azoto e ossigeno; altri elementi pesanti, quali neon, zolfo, sodio, argo e cloro, sono presenti in concentrazioni ancora inferiori.

Si ritiene che molti di questi elementi vengano sintetizzati nelle reazioni termonucleari che si verificano nelle regioni centrali delle stelle; la loro presenza in una nebulosa induce quindi a ritenere che siano stati in qualche modo «estratti» dal nucleo della stella. Le nebulose planetarie non presentano tracce di fenomeni violenti, quindi la materia deve essere stata trasportata per convezione, il processo turbolento che trasferisce materia ed energia attraverso la stella. Dall'abbondanza relativa degli elementi pesanti nelle nebulose planetarie si ricava per esempio che la materia che le costituisce deve essere stata estratta dal nucleo ricco di elio e trasportata verso gli strati esterni della stella; il forte vento emesso dall'astro ha poi disperso questi strati. L'analisi della composizione delle nebulose planetarie rivela alcuni parametri indicativi dei processi verificatisi nella stella progenitrice, soprattutto l'efficienza e la profondità della convezione al suo interno.

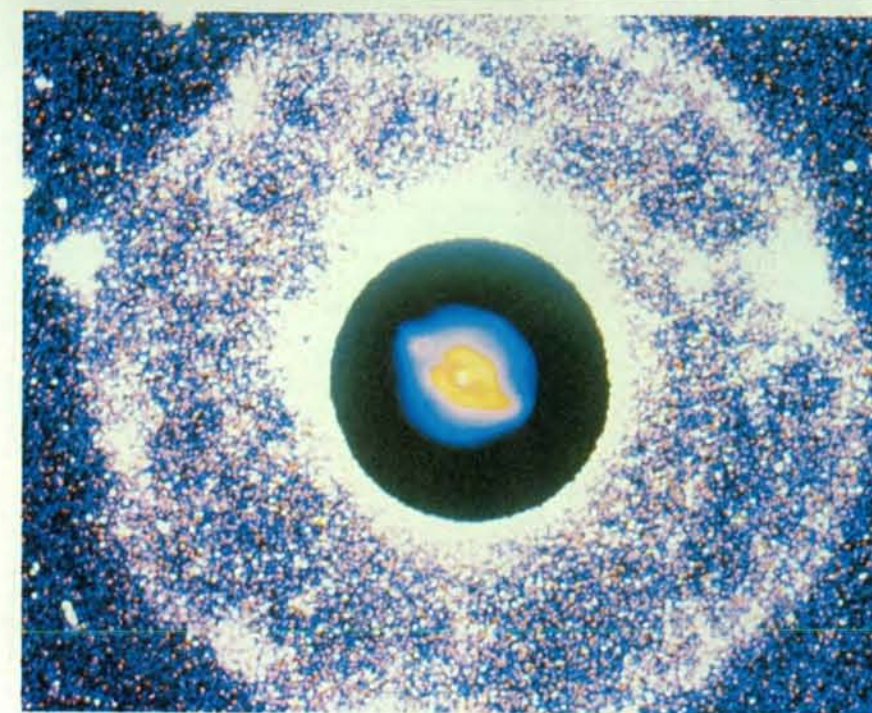
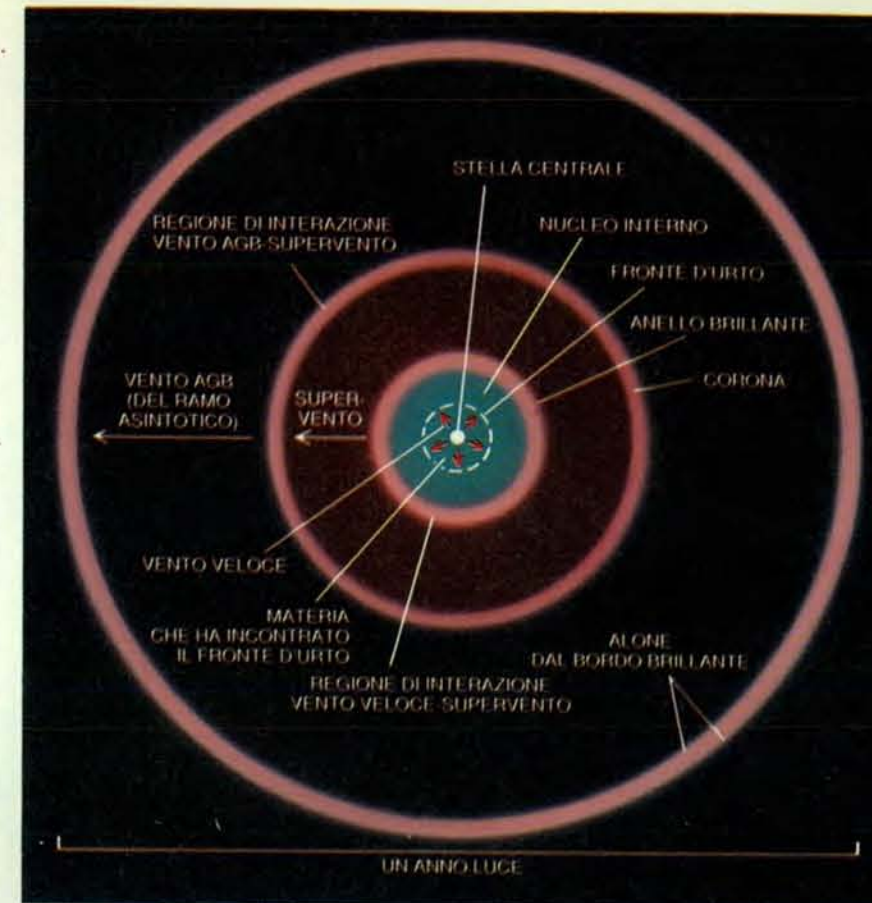
L'intensità relativa delle varie righe spettrali rivela anche la temperatura e la densità delle nebulose planetarie. Si ricava che la temperatura media è di norma intorno ai 10 000 kelvin, mentre la densità varia da alcuni atomi per centimetro cubo al centro e alla periferia estrema fino a ben 100 000 atomi per centimetro cubo nelle regioni del bordo più brillanti e compresse. Densità simili sono migliaia di volte maggiori di quella della materia interstellare. (Le nebulose planetarie restano comunque decisamente «vuote» secondo i criteri terrestri: un centimetro cubo di aria al livello del mare contiene circa 10^{19} atomi.) Le nebulose più giovani, e quindi più piccole, tendono ad avere densità relativamente elevate.

Le numerose scoperte sulla fisica delle nebulose planetarie hanno fornito ai teorici un valido contributo nell'elaborazione di idee più precise sull'evoluzione stellare. Negli ultimi 20 anni Bohdan Paczynski, ora alla Princeton University, e poi Icko Iben, Jr., dell'University dell'Illinois, nonché Detlef Schönberner dell'Università di Kiel e molti altri hanno realizzato simulazioni al calcolatore assai complesse di modelli stellari, che hanno permesso agli astronomi di ricostruire un quadro delle varie fasi dell'evoluzione stellare fino alla formazione delle nebulose planetarie.

Approssimativamente per il 90 per cento della propria vita, le stelle producono regolarmente energia nella regione centrale tramite fusione nucleare: nuclei di idrogeno si fondono formando nuclei di elio e nel processo si libera energia. In questa fase stabile la stella appartiene alla cosiddetta sequenza principale, della quale, fortunatamente per la vita sulla Terra, fa parte anche il Sole. La durata della permanenza di una stella nella sequenza principale dipende molto dalla sua massa: le stelle pesanti splendono con intensità molto maggiore e quindi consumano il proprio idrogeno assai più velocemente di quelle leggere. Perciò, mentre una stella di massa prossima a quella solare resta nella sequenza principale per circa 10 miliardi di anni, una stella di massa cinque volte maggiore vi resiste solo per 60 milioni di anni.

Via via che una stella invecchia, nel suo nucleo si accumula elio. Quando l'idrogeno qui contenuto si esaurisce completamente, il nucleo si contrae sotto la propria attrazione gravitazionale e si riscalda, mentre la fusione si trasferisce più all'esterno, in un guscio, che circonda il nucleo, dove esiste ancora materia ricca di idrogeno. Nel corso di questa fase la stella si espande fino a un diametro pari a decine di volte quello del Sole (che è di circa 1,4 milioni di chilometri). La superficie dell'astro, ormai molto lontana dalla fonte centrale di energia, si raffredda fino a divenire rossa: a causa dell'aumento di superficie, però, la luminosità totale aumenta comunque di oltre 100 volte. In questa fase della propria evoluzione la stella prende il nome di gigante rossa.

La fusione dell'idrogeno nel guscio esterno continua a far accumulare «ceneri» di elio sul nucleo, la cui massa e temperatura aumentano sempre più. Quando la temperatura del centro della stella arriva intorno a 100 milioni di kelvin i nuclei di elio cominciano anch'essi a subire fusione, con formazione di nuclei di carbonio e ossigeno. Nel frattempo le zone esterne della stella si contraggono facendo riscaldare la superficie, che assume un colore più blu. Alla fine anche l'elio nel centro si esaurisce; via via che si inaridisce la fonte centrale di energia, le forze gravitazionali tornano a prendere il sopravvento e il nucleo si contrae riscaldandosi e addensandosi. Ora anche la fusione dell'elio si trasferisce in un guscio più esterno; la stella assume perciò una struttura a cipolla, con un guscio esterno, dove si ha la fusione dell'idrogeno, e uno più interno, dove avviene la fusione dell'elio, che circondano un nucleo inerte di carbonio e ossigeno.

Dopo la formazione del guscio di elio, la stella si espande fino a raggiungere un

Lo schema di una nebulosa planetaria (*in alto*) mostra una regione interna di gas perturbato da un vento veloce emesso dalla stella centrale. Dove questo vento collide col supervento, più denso ma meno veloce, si sviluppa un anello brillante. A distanze maggiori anche il supervento interagisce con materia liberata in una fase precedente della vita della stella, dando origine a una corona. Il vento più antico (vento AGB), che ha avuto più tempo per allontanarsi dalla stella, può comprimere la materia interstellare circostante dando origine a un alone dal bordo brillante. Un'immagine di NGC 6826 (*in basso*) rivela molti di questi caratteri. La sensibilità delle regioni esterne dell'immagine è stata aumentata di molto per rivelare i dettagli più deboli.

Lo spettro ultravioletto della nebulosa planetaria NGC 6826 registra l'emissione di radiazione da parte di ioni del vento veloce (*picchi*) e il suo assorbimento da parte di ioni in moto verso la Terra proiettati sulla stella centrale (*minimi*). La lunghezza d'onda e l'intensità delle righe indicano che il vento si muove a 2000 chilometri al secondo e sottrae alla stella circa un decimilionesimo di massa solare all'anno.

diametro centinaia di volte maggiore di quello del Sole ed entra in quella regione del diagramma stellare (o diagramma di Hertzsprung-Russell) detta ramo asintotico delle giganti. In seguito a un'espansione così intensa la temperatura superficiale precipita a circa 3000 kelvin (la temperatura alla superficie del Sole, per confronto, è di circa 6000 kelvin).

La successione esatta di eventi della fase che segue non è ancora perfettamente chiara perché l'evoluzione avviene con grande rapidità. Una gigante del ramo asintotico produce un vento stellare molto intenso che ne espelle lo strato esterno. L'esistenza di questi venti è attestata chiaramente, ma il meccanismo che li genera è ancora piuttosto misterioso. All'elevata velocità con cui la stella perde massa è dovuto probabilmente il fatto che le giganti del ramo asintotico vanno incontro a fluttuazioni significative di raggio, temperatura e luminosità. Queste stelle sono classificate fra le variabili del tipo Mira, che prendono il nome dalla stella pulsante molto luminosa Mira Ceti, assunta a paradigma della categoria.

Anche le stelle che si trovano nella sequenza principale espellono una moderata quantità di materia, ma una stella come il Sole perde solo un decibilionesimo di massa solare all'anno, sotto forma di vento che si allontana a una velocità di circa 600 chilometri al secondo. Quando la stella diventa una gigante del ramo asintotico la perdita di massa sale a circa un milionesimo di massa solare all'anno.

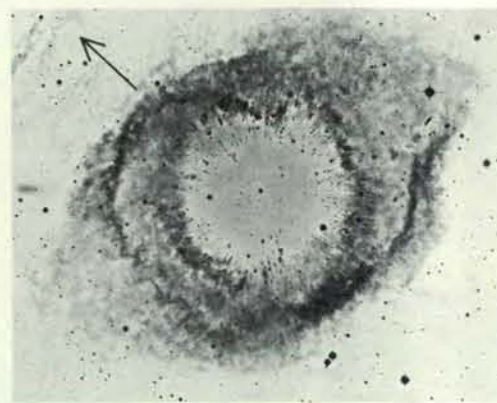
Nelle fasi successive queste stelle emettono un flusso ancora più violento, il cosiddetto supervento, che trascina con sé anche un decimillesimo di massa solare all'anno e consuma quasi completamente i resti dell'involucro esterno della stella. Le giganti del ramo asintotico sono tanto grandi che il legame gravitazionale che trattiene i loro strati esterni è molto debole, sicché il gas non ha bisogno di raggiungere una velocità particolarmente elevata per sfuggire. Il supervento si allontana quindi all'andatura moderata di 10 o 20 chilometri al secondo.

Intorno all'inizio degli anni ottanta numerosi teorici, tra i quali Amos Harpaz e Attay Kovetz dell'Università di Tel Aviv, hanno delineato un modello quantitativo dettagliato della trasformazione delle stelle giganti in nebulose planetarie. Dopo che la gigante si è liberata di quasi tutto il proprio involucro esterno il supervento tende ad affievolirsi: tutto ciò che resta della stella originaria è il nucleo rovente, completamente spogliato. Questo, a sua volta, si contrae e diventa sempre più caldo, alimentato dalla fusione di quel poco di idrogeno che ancora rimane. A onta di tutte queste trasformazioni la luminosità della stella resta costante, pari a migliaia di volte quella del Sole. Una circostanza notevole è che a questo punto la sua massa è uguale a circa 0,6 volte quella del Sole, indipendentemente dalla massa iniziale dell'oggetto quando si trovava nella sequenza principale.

La nube di gas intorno alla stella si raffredda progressivamente durante l'espansione. Quando la temperatura scende al di sotto dei 1000 kelvin circa cominciano ad addensarsi granuli di polvere composti da carbonio, ossigeno, silicio e altri elementi pesanti, ciascuno di diametro compreso tra un decimillesimo e un millesimo di millimetro. La polvere nasconde la stella assorbendone quasi tutta la radiazione visibile e riemettendola come radiazione infrarossa, di minore energia e lunghezza d'onda maggiore. Si sono osservate queste protonebulose planetarie, invisibili ai telescopi ottici, con rivelatori infrarossi a terra e, nel 1983, con l'*Infrared Astronomical Satellite* in orbita intorno alla Terra.

Diverse centinaia o migliaia di anni dopo la fine del supervento la superficie della stella centrale in via di contrazione si è riscaldata fino a una temperatura di 25 000 kelvin. A questo punto la radiazione emessa dalla stella ha un'energia così elevata da strappare elettroni agli atomi della nebulosa; ecco perché quasi tutto il gas che si osserva nelle nebulose planetarie è ionizzato. Collidendo con gli elettroni liberati o combinandosi con essi, gli ioni emettono righe spettrali intense, soprattutto alle lunghezze d'onda del visibile e dell'ultravioletto.

La radiazione di alta energia della stella viene assorbita innanzitutto dalle regioni interne della nebulosa. Via via che la stella centrale continua a riscaldarsi e la nebulosa a espandersi (diventando perciò sempre meno densa), si ionizzano regioni sempre più ampie della nebulosa. Nello stesso tempo la polvere si dirada e la luminosità della nebulosa nella regione visibile dello spettro aumenta rapidamente. Quando la nube di gas diventa abbastanza brillante da poter essere osservata tramite telescopi ottici a terra, l'oggetto viene classificato tra le nebulose planetarie. Al culmine della sua attività, qualche migliaio di anni dopo l'espulsione della nebulosa planetaria, la stella centrale raggiunge una temperatura di 200 000 kelvin e una luminosità simile a quella che aveva quando era una gigante (ossia varie migliaia di volte quella del Sole).

È la penetrazione progressiva della radiazione ionizzante attraverso la nebulosa a spiegare la correlazione positiva che si osserva tra le dimensioni di una nebulosa planetaria e la sua massa ionizzata totale. Le nebulose planetarie più leggere hanno una massa pari a circa 0,01 volte quella del Sole e un raggio di circa 0,05 anni luce, mentre quelle di massa più grande contengono circa una massa solare e si estendono in un raggio di diversi anni luce. La lunghezza d'onda e l'intensità della radiazione emessa dai vari elementi dipendono da parametri quali la temperatura, la densità e l'energia della radiazione ionizzante, che variano notevolmente da un punto all'altro della nebulosa.

Una modesta quantità di materia continua ad allontanarsi dalla stella centrale anche dopo la fase di gigante del ramo asintotico. A questo punto la stella perde massa lentamente (da un decimiliardesimo a un decimilionesimo di massa solare all'anno), ma questo flusso così debole di materia si muove a una velocità estremamente elevata, compresa tra 1000 e 4000 chilometri al secondo. Si tratta del vento veloce che si osserva studiando lo spettro di alcune nebulose planetarie. La sua elevata velocità è dovuta al fatto che adesso la stella è molto più piccola e densa, e quindi la materia deve muoversi circa 100 volte più velocemente del supervento per sfuggire alla gravità che si esercita alla superficie stellare.

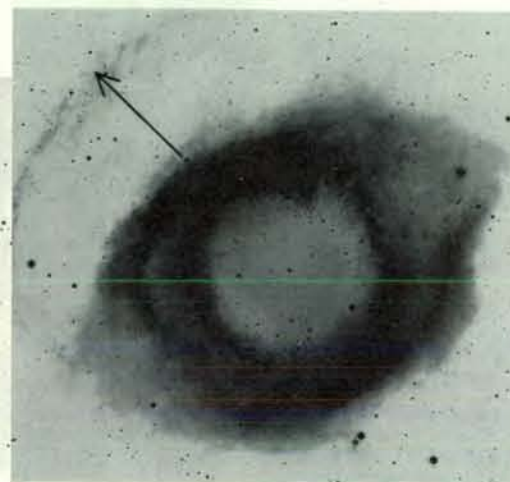
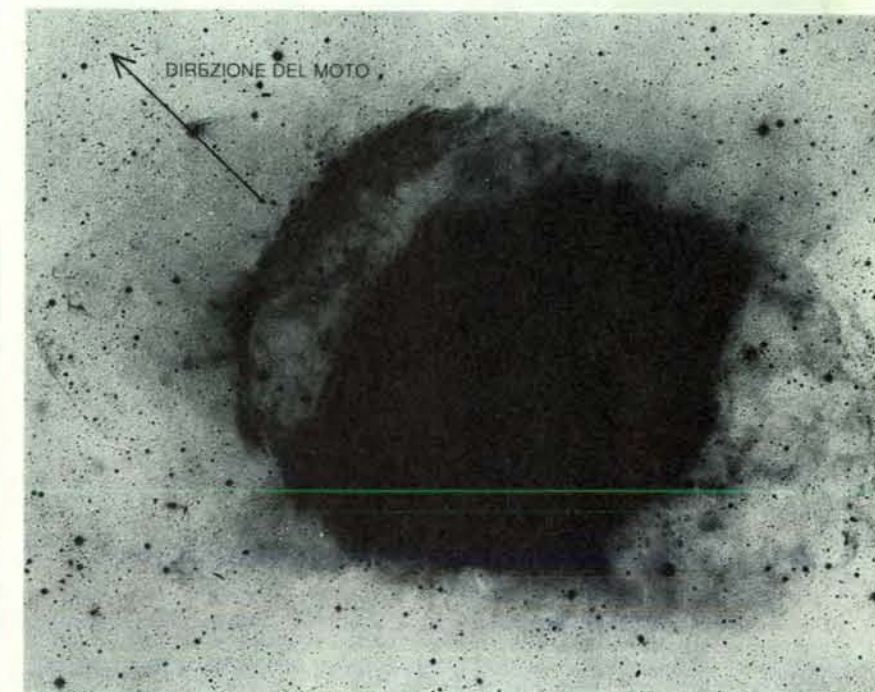
Anche se la massa totale che trasporta è trascurabile, questo vento riveste un ruolo di prim'ordine nello sviluppo delle nebulose planetarie. I primi ad avanzare l'ipotesi che il vento veloce abbia importanza a questo riguardo sono stati Sun Kwok dell'Università di Calgary, C. R. Purton del Dominion Radio Astrophysical Observatory e M. Pim FitzGerald dell'Università di Waterloo. Il vento veloce forma un'onda d'urto circondata da una bolla di gas rarefatto molto caldo nella regione centrale della nebulosa. La pressione della bolla accelera la materia più esterna trasportata dal supervento, che si muove più lentamente; dove il secondo flusso di gas raggiunge il primo si forma un guscio sottile di gas particolarmente denso, che assume l'aspetto di un anello brillante.

In alcuni casi il fronte esterno del supervento comprime la materia trasportata dal vento che la stella aveva emesso in precedenza, quando era una gigante del ramo asintotico; il fenomeno può dare origine a un altro guscio brillante, la cosiddetta corona. Alla stessa stregua, anche questo vento (vento AGB, o del ramo asintotico delle giganti) può interagire con la materia interstellare, ancor più rarefatta, producendo un debole alone dal bordo luminoso.

Dopo varie migliaia di anni la stella diventa estremamente compatta e il vento veloce cessa. Esaurito l'idrogeno residuo vicino alla superficie, la fusione nucleare termina e la stella inizia a raffreddarsi e a oscurarsi mentre anche la nebulosa planetaria, espandendosi, riduce la propria luminosità per unità di superficie. Alla fine la nebulosa scompare completamente e la stella centrale, compatta e nuda, diventa una nana bianca. La durata del vento veloce e il tempo necessario per bruciare tutto l'idrogeno dipendono notevolmente dalla massa iniziale della stella centrale: le stelle di grande massa splendono di più e si evolvono più rapidamente di quelle leggere. Da 10 000 a 30 000 anni circa dopo essere comparsa la nebulosa planetaria diventa invisibile.

Non tutte le stelle attraversano una fase di nebulosa planetaria. Una massa inferiore a 0,8 volte quella del Sole è insufficiente per dare origine a una nebulosa, ma anche le stelle di massa superiore a otto masse solari sono escluse perché muoiono in una violentissima esplosione di supernova. Se ne ricava che solo le stelle di massa compresa tra 0,8 e 8 volte quella del Sole possono generare una nebulosa planetaria.

A conferma di questa valutazione giunge la distribuzione delle nebulose planetarie nella nostra galassia. Quasi tutte le nebulose conosciute si concentrano nel piano del disco della Via Lattea, dove le stelle sono in grande maggioranza giovani. Questa predominanza indica che le nebulose planetarie si sono evolute per lo più a partire da stelle giovani e a breve vita, di massa maggiore di quella del Sole. Al contrario, nelle regioni più esterne dell'alone galattico - la sfera poco densa di stelle che circonda il disco galattico - si sono osservate solo cinque nebulose planetarie. Le stelle dell'alone hanno un'età media di 10 miliardi di anni, il che induce a ritenere che queste nebulose planetarie debbano essersi evolute da stelle di massa uguale o leggermente inferiore a quella solare.

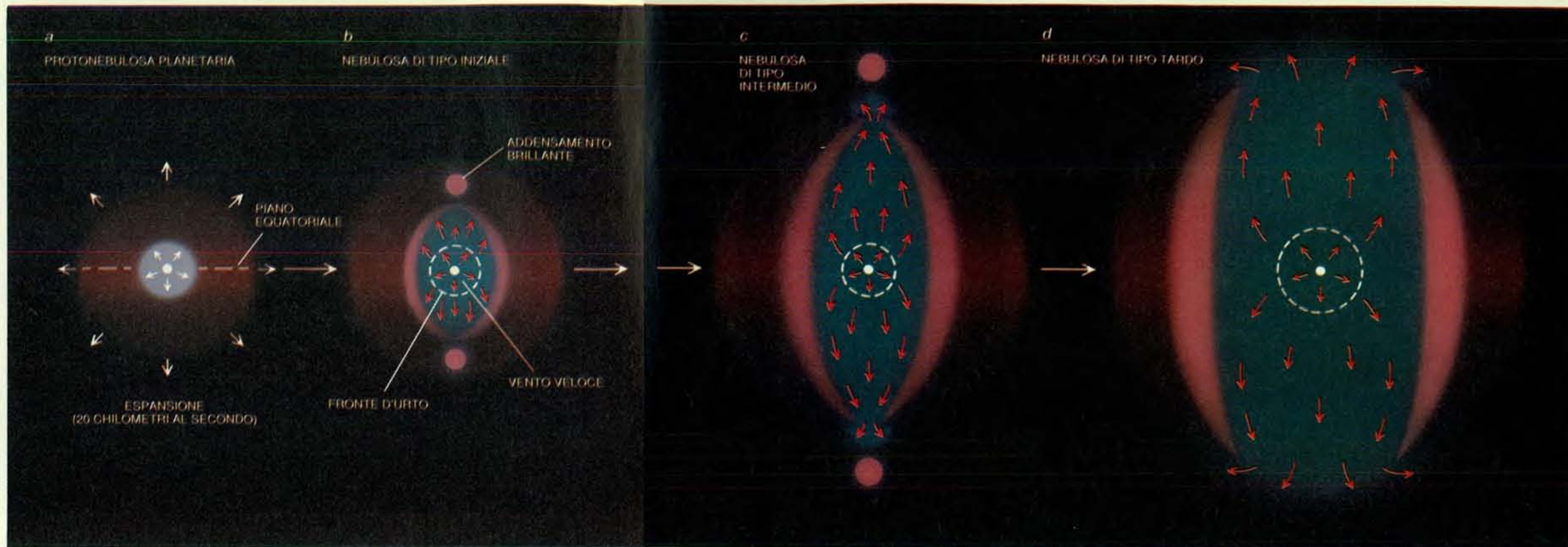
La struttura di una nebulosa planetaria dipende molto dall'ambiente ove essa nasce e si sviluppa. In assenza di influenze esterne dovrebbero formarsi bolle di gas perfettamente sferiche, ma meno di un decimo delle nebulose planeta-

Questa serie di immagini rivela la complessa struttura della nebulosa Elica. La regione ellittica maculata più interna (*a sinistra*) rispecchia l'irregolarità del flusso di materia espulso dalla stella centrale. Il moto della stella attraverso il mezzo interstellare comprime un arco di gas più esterno, come si vede nella parte in alto a sinistra di una immagine ottenuta con tempo di esposizione più lungo (*al centro*). L'immagine realizzata con la massima sensibilità (*a destra*) sovraespone le regioni centrali, ma permette di individuare una formazione conica poco luminosa davanti alla nebulosa e ondulazioni del gas a bassa densità dietro di essa.

rie catalogate ha questa forma: quasi tutte presentano un evidente asse di simmetria, che rivela l'azione di altre forze.

Le nebulose planetarie con un'asimmetria modesta come NGC 3242 si dicono ellittiche; quelle ad asimmetria marcata come NGC 650-1 (il Piccolo manubrio) e NGC 2346 vengono definite a farfalla. Molte nebulose contengono anche due addensamenti brillanti ai poli opposti della stella: è il caso di NGC 7009. Tutte queste nebulose possiedono una simmetria assiale, ovvero constano di due metà simmetriche rispetto a un asse centrale.

Secondo quasi tutti i modelli teorici, le forme a simmetria assiale sarebbero il risultato di una perdita di massa particolarmente rilevante nel piano equatoriale della stella durante la fase di supervento. Bruce Balick dell'Università di Washington ha sottolineato che i dati osservativi, secondo i quali il gas si concentra nel piano equatoriale delle stelle, confermano nel complesso questi modelli.

Mark R. Morris dell'Università della California a Los Angeles ha avanzato l'ipotesi che le nebulose planetarie non sferiche possano svilupparsi intorno a sistemi stellari doppi. Secondo la sua teoria, durante la fase di gigante del ramo asintotico la stella progenitrice crescerebbe fino a far traboccare il proprio involucro esterno sulla compagna, dando origine a un disco di gas in rotazione intorno a quest'ultima. Si formerebbero allora due getti di materia che si estendono in direzione perpendicolare al disco (il meccanismo fisico alla base di quest'ultimo fenomeno è tutt'altro che chiaro, ma si osservano spesso getti espulsi dai poli di stelle circondate da un disco di materia). Sarebbero gli effetti mareali della compagna a far concentrare il supervento nel piano orbitale del sistema. La materia espulsa nei getti polari formerebbe poi addensamenti luminosi di gas allineati in direzione perpendicolare al disco.

In collaborazione con Mario Livio del Technion-Israel Institute of Technology, ho elaborato un modello che spiega l'origine delle nebulose planetarie a ellitticità meno marcata. Il nostro modello è analogo a quello sviluppato da Morris, a parte il fatto che la compagna orbita più vicino alla stella progenitrice, all'interno del suo involucro. Le forze mareali e l'attrazione gravitazionale fanno diminuire progressivamente il diametro dell'orbita della compagna e, per compensazione, il gas nel piano orbitale di quest'ultima si muove più velocemente verso l'esterno, facendo sì che il supervento risulti particolarmente intenso nel piano dell'orbita.

In questo caso la nebulosa, espandendosi, assume una forma ellittica anziché sferica. Si è riscontrato che diverse nebulose ellittiche e a farfalla hanno effettivamente al centro sistemi binari con un periodo orbitale che va da qualche ora a qualche giorno. La brevità del periodo indica che le due stelle orbitano molto vicine, a pochi milioni di chilometri l'una dall'altra.

La molteplicità di forme delle nebulose planetarie ha una spiegazione: il flusso di materia emesso dalla stella si concentra spesso nel piano equatoriale (*a*), probabilmente a causa degli effetti gravitazionali di un'altra stella, che forma un sistema doppio con la prima. La nebulosa assume perciò una forma ellittica o a farfalla. La stella può anche emettere gas lungo l'asse polare, creando addensamenti brillanti (*b*). Il vento veloce (*frecce rosse*) proveniente dalla stella genera un guscio di gas compresso dove incontra il supervento. Il vento veloce esce prima ai poli (*c*) e poi si espande nel mezzo interstellare (*d*).

La forma di una nebulosa planetaria può essere influenzata anche dal suo moto attraverso il mezzo interstellare. Quest'ultimo è estremamente rarefatto: la sua densità è di circa un atomo per centimetro cubo nel piano galattico e scende bruscamente subito al di sopra e al di sotto di esso. Mentre una stella e la nebulosa che la circonda si muovono nella Via Lattea, il mezzo interstellare le lambisce a una velocità compresa tra 40 e 150 chilometri al secondo. La parte della nebulosa che fronteggia il flusso rallenta e viene compressa nella direzione del moto; diventa quindi relativamente più densa e più brillante della materia circostante.

Laddove la materia della nebulosa incontra il mezzo interstellare possono svilupparsi formazioni complesse. Nel caso della nebulosa Elica, per esempio, sembra che un gas molto rarefatto sia stato rimescolato e compresso un poco più avanti della nebulosa, lungo la direzione del moto. Attualmente si conosce soltanto una decina di nebulose planetarie che presentino tracce evidenti di interazione col mezzo interstellare. Se si riuscisse ad ampliare questo esiguo campione in misura significativa, gli astronomi avrebbero a disposizione uno strumento prezioso per studiare la densità e altre caratteristiche fisiche del mezzo interstellare.

A onta dei numerosi progressi compiuti nello studio delle nebulose planetarie, alcuni dei parametri fisici fondamentali di questi oggetti, come la massa e le dimensioni, non sono ancora determinati con precisione, a causa della difficoltà di valutarne la distanza. Per fortuna parecchie nebulose fanno parte di sistemi stellari molto ben conosciuti. Due di esse sono state osservate in ammassi globulari - ammassi sferici densi che contengono fino a un milione di stelle - dei quali si conosce la distanza con grande precisione. Sembra inoltre che varie centinaia di nebulose si trovino vicino al centro galattico, sicché è legittimo supporre che la loro distanza dalla Terra sia all'incirca la stessa alla quale giace il centro galattico, ossia circa 25 000 anni luce.

Anche le stelle di ciascuna galassia esterna si trovano tutte, in buona approssimazione, alla stessa distanza dalla Terra. La distanza delle nebulose planetarie nelle galassie vicine può quindi essere valutata con discreta precisione. Nella Piccola Nube di Magellano (una delle galassie satelliti della Via Lattea, a circa 190 000 anni luce di distanza) si è individuata una sessantina di questi oggetti, 130 nella Grande Nube di Magellano (a 160 000 anni luce), e varie centinaia nella galassia di Andromeda (lontana due milioni di anni luce).

Per quasi tutte le nebulose planetarie le tecniche attuali di misurazione delle distanze non riescono a ridurre il margine d'errore al di sotto del 50 per cento in più o in meno. Una di queste tecniche, ideata da Julie H. Lutz della Washington State University, si basa sull'indebolimento della radiazione emessa dall'oggetto a causa della polvere interstellare, un fenomeno chiamato estinzione. Nel visibile e nell'ultravioletto l'estinzione aumenta al diminuire della lunghezza d'onda. Osservando il rapporto tra l'intensità di due righe in emissione a lunghezze d'onda diverse e confrontandolo con quello previsto dalla teoria, la Lutz è riuscita quindi a ricavare l'entità totale dell'affievolimento. È poi riuscita a calcolare la distanza dalla valutazione dell'estinzione per unità di distanza in direzione della nebulosa (stimata misurando la luce delle stelle presenti nella stessa regione del cielo).

Diversi metodi di determinazione delle distanze si affidano ai valori medi di certe grandezze calcolati per un gran numero di nebulose planetarie. Se si ipotizza per esempio che queste contengano tutte la stessa quantità di gas ionizzato, si può estrapolare una distanza sconosciuta a partire da una nota. Per determinare la distanza di una nebulosa planetaria se ne confrontano l'estensione angolare e l'intensità dell'emissione dell'idrogeno ionizzato con quelle di un'altra nebulosa la cui distanza è già nota. Questo metodo è estremamente impreciso per una specifica nebulosa, ma può fornire valori medi accettabili per un gruppo numeroso.

Le misurazioni di distanze basate su campioni di nebulose planetarie hanno in potenza applicazioni cosmologiche molto importanti. Si sta perseguendo attivamente la ricerca di tecniche di misurazione affidabili della distanza delle galassie lontane per stabilire il valore della velocità di espansione dell'universo, la cosiddetta costante di Hubble. La sua determinazione permetterebbe di ricavare le dimensioni e l'età dell'universo stesso, informazione che sarebbe a sua volta utile per risolvere vari problemi sulle modalità e l'epoca dell'inizio e della fine del cosmo. Attualmente la costante di Hubble è nota solo con l'approssimazione di un fattore due.

In genere si misura la distanza delle galassie vicine studiando una categoria di stelle, le variabili Cefeidi. Le nebulose planetarie sono brillanti e facili da identificare anche in galassie nelle quali le singole stelle non sono distinguibili. Inoltre queste nebulose emettono intensa radiazione a lunghezze d'onda ben precise, il che rende più facile distinguerne l'emissione da quella del resto della galassia in cui si trovano. Robin Ciardullo della Pennsylvania State University, George Jacoby del Kitt Peak National Observatory e Holland Ford dello Space Telescope Science Institute, con vari collaboratori, hanno perciò tentato di utilizzare le nebulose planetarie come indicatori delle distanze celesti.

Questi studiosi partono dal presupposto che tipi di galassie simili contengano popolazioni simili di nebulose planetarie. In particolare, dividendo le nebulose in diverse classi di luminosità, la percentuale della popolazione che cade in ciascuna classe dovrebbe rimanere pressoché costante da galassia a galassia. Dal momento che l'obiettivo è determinare la distanza delle galassie lontane, gli astronomi si sono concentrati sulle nebulose planetarie più brillanti, osservandole in particolare alla lunghezza d'onda della riga dell'ossigeno ionizzato due volte (a 500,7 nanometri), dove è facile identificarle.

Ciardullo e colleghi hanno riscontrato che, anche in galassie la cui struttura e composizione stellare differiscono leggermente, le nebulose planetarie più brillanti hanno effettivamente circa la stessa luminosità a 500,7 nanometri e ne hanno dedotto che gli oggetti di questo

sottogruppo possano essere utilizzati come «candele campione», cioè come sorgenti di luminosità assoluta nota e indipendente dalla posizione.

Dalla luminosità prevista delle nebulose più brillanti si può ricavare la distanza di qualunque galassia dove sia possibile identificare con sicurezza le nebulose planetarie. Stuart R. Pottasch del Kapteyn Astronomical Institute nei Paesi Bassi ha dimostrato che questo metodo fornisce valutazioni precise della distanza del centro galattico, già misurata con altre tecniche più consolidate.

Il gruppo di Ciardullo ha quindi valutato la distanza delle galassie dell'ammasso della Vergine ricavandone un valore della costante di Hubble che si colloca verso l'estremità più alta dell'intervallo di valori stimati. Questi risultati comportano che l'universo sia più piccolo e giovane (circa 10 miliardi di anni di età) di quanto non pensi la maggior parte dei cosmologi, ma si tratta di valori provvisori e abbastanza discussi. Osservazioni di nebulose planetarie in altre galassie e confronti con i risultati dell'impiego di altri indicatori della distanza aiuteranno senza dubbio a raffinare la tecnica che utilizza le nebulose planetarie come candele campione.

Molte altre caratteristiche delle nebulose planetarie reclamano a gran voce ulteriori studi: il meccanismo che regola la rapidissima trasformazione delle stelle giganti del ramo asintotico in nebulose è ancora poco chiaro, e anche la natura del supervento e del vento veloce e le loro modalità di interazione rimangono un mistero. Raffinando progressivamente gli studi sulle nebulose planetarie, gli scienziati scopriranno nuove informazioni su una gamma di ambienti spaziali che va dalle regioni interne di vecchie stelle simili al Sole alla materia rarefatta che pervade tutta la Via Lattea, e riusciranno anche a comprendere sempre meglio quel fugace momento di gloria finale che attende il Sole tra cinque miliardi di anni.

BIBLIOGRAFIA

POTTASCH STUART R., *Planetary Nebulae*, D. Reidel Publishing, 1984.

KALER JAMES B., *Planetary Nebulae and Their Central Stars* in «Annual Review of Astronomy and Astrophysics», 23, 1985.

KALER JAMES B., *Stars and Their Spectra: An Introduction to the Spectral Sequence*, Cambridge University Press, 1989.

PLAIT PHILIP e SOKER NOAM, *The Evolution of the Planetary Nebula NGC 6826* in «Astronomical Journal», 99, n. 6, giugno 1990.

SOKER NOAM, *On the Formation of Ansae in Planetary Nebulae* in «Astronomical Journal», 99, n. 6, giugno 1990.

Il ruolo dell'Antartide nell'evoluzione del Gondwana

Situata al centro dell'antico supercontinente, l'Antartide ha avuto nella storia del Gondwana una parte determinante che acquista una fisionomia sempre più definita al crescere delle conoscenze geologiche

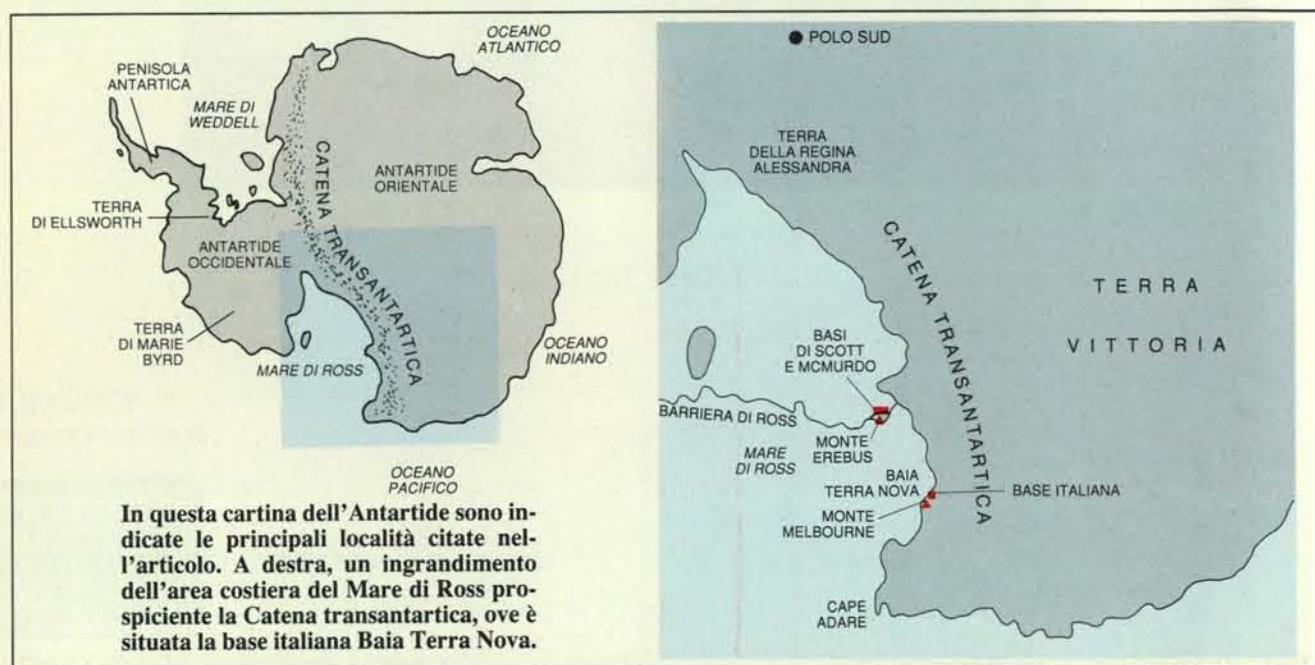
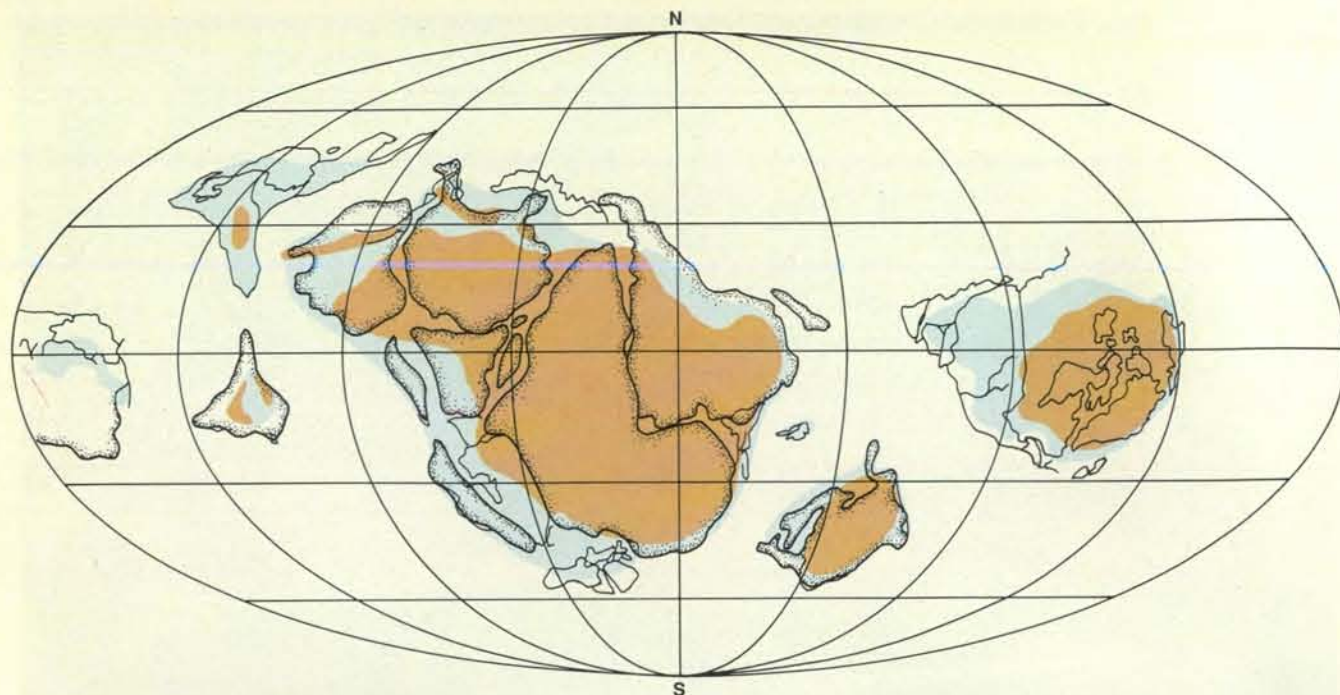
di Raffaele Casnedi e Bruno Lombardo

In una fase della storia della Terra, tutti i continenti erano riuniti in un'unica grande massa. Su questa asserzione concordano le teorie geodinamiche dalla loro concezione più classica, rappresentata dalla teoria della deriva dei continenti di Alfred Wegener, fino alla moderna affermazione della teoria della tettonica a zolle. Questa fase è rappresentata dal periodo Carbonifero, che offre interessanti dati paleogeografici, primi fra tutti quelli paleoclimatici, testimoniati dallo sviluppo di una ricca flora. Questo blocco continentale, denominato Pangea, si sarebbe in seguito frammentato negli attuali continenti. Il loro movimento e la posizione reciproca sono delineati in ampie ricostruzioni paleogeografiche che diventano via via più dettagliate con l'evolversi delle conoscenze scientifiche.

Molto scarse sono viceversa le conoscenze geologiche sulle fasi di deriva continentale che precedono il Carbonifero, ed è anche su questo tema che lo studio geologico dell'Antartide può offrire nuovi elementi per interpretare i movimenti che si sono verificati nel Paleozoico inferiore, e cioè fra 600 e 350 milioni di anni fa.

All'Antartide erano infatti anticamente collegati tutti i continenti che attualmente sono situati nell'emisfero australe. In essa sono largamente rappresentati in affioramento (ove non siano ricoperti dalla spessa coltre glaciale) terreni - come intrusioni magmatiche e sedimenti - che si sono formati nelle fasi precambriane e del Paleozoico inferiore. Inoltre, in gran parte dell'Antartide, questi terreni non hanno subito deformazioni successive.

Già nel secolo scorso, in iceberg alla deriva verso il Sud America furono prelevati frammenti di rocce provenienti dall'Antartide. Il primo campione *in situ* fu raccolto nel 1898 dalla spedizione Borchgrevink a Cape Adare, all'estremità del continente antartico prospiciente la Nuova Zelanda. Fu in questa zona, e all'estremità opposta, cioè nella Penisola antartica (più facilmente accessibile in quanto situata non lontano dall'America Meridionale), che fu avviata nel «conti-

In questa cartina dell'Antartide sono indicate le principali località citate nell'articolo. A destra, un ingrandimento dell'area costiera del Mare di Ross prospiciente la Catena transantartica, ove è situata la base italiana Baia Terra Nova.

In alto, il cono vulcanico del Monte Melbourne che domina la banchisa polare in prossimità della base italiana Baia Terra Nova. In basso, Cape Adare, all'estremità dell'Antartide sul versante del Pacifico, il punto di sbarco della spedizione di Borchgrevink del 1898. Le capanne sono quelle che ripararono questi primi esploratori dai rigori dell'inverno australe.

Questa ricostruzione mostra una ipotetica distribuzione dei continenti e degli oceani nel Cambriano (oltre 500 milioni di anni fa). Il supercontinente Gondwana è già definito geograficamente e in esso sono riportati i margini degli attuali continenti. L'Antartide - e in particolare la zona della Terra Vittoria, che corrisponde qui al bordo meridionale del continente - è situata in prossimità dell'equatore, e quindi si spiega perché vi siano state rinvenute faune cambriane di clima caldo. Si noti la posizione invertita dei continenti rispetto a quella attuale: essi non avevano ancora iniziato il moto di deriva in blocco che li avrebbe portati ad attraversare il Polo sud e poi a risalire verso nord nella posizione che hanno oggi.

nente di ghiaccio» la ricerca geologica.

All'inizio del nostro secolo queste aree furono raggiunte da molte spedizioni di diverse nazionalità con finalità esplorative, che limitarono però le loro osservazioni alla fascia costiera, salvo le storiche eccezioni delle spedizioni polari di Robert Scott, Ernest Shackleton e Roald Amundsen. La ricerca geologica ebbe tuttavia il suo maggiore sviluppo a partire dall'Anno geofisico internazionale del 1957-1958.

L'esplorazione della Terra Vittoria

Il contributo italiano all'esplorazione scientifica dell'Antartide è attualmente rappresentato dal Programma nazionale di ricerca scientifica e tecnologica in Antartide (PNRA) iniziato nel 1985, che impegna ricercatori delle università, del Consiglio nazionale delle ricerche e dell'ENEA. Esso ha avuto come punto di riferimento la base di Baia Terra Nova, sulla costa del Mare di Ross (Terra Vittoria), in prossimità del vulcano Monte Melbourne. Scoperto da Sir James Clark Ross - ammiraglio della spedizione inviata dal Governo britannico nel 1841 alla ricerca del polo sud magnetico - che battezzò col nome della giovane regina la terra antistante, il Mare di Ross costituì il punto di partenza per le osservazioni geologiche condotte da John King David e Sir Douglas Mawson della spedizione Shackleton (1907-1909), mentre il Gruppo nord della seconda spedizione Scott (1910-1913), che esplorò i dintorni della Baia Terra Nova prima di esservi costretto a svernare in condizioni estreme, includeva il geologo Raimond Edward Priestley.

Dopo questa fase pionieristica, l'esplorazione geologica della Terra Vittoria è proseguita a opera di ricercatori neozelandesi e statunitensi. Dopo l'Anno geofisico internazionale essi hanno organizzato quasi ogni estate australe

Sono riportati i maggiori eventi ignei, sedimentari e metamorfici che hanno interessato una parte della Catena transantartica fino alla frammentazione del Gondwana. Sul basamento cristallino precambriano sedimentano calcari ad archeociatidi e conglomerati (sulla destra della figura, che corrisponde agli affioramenti nei Monti della Regina Alessandra). Ai calcari seguono in probabile sovrapposizione (che non è stata provata poiché non esiste continuità di affioramento fra le due aree) depositi detritici torbiditici che figurano sulla sinistra e che corrispondono alla Terra Vittoria. Nell'Ordoviciano, l'orogenesi di Ross si manifesta con deformazione, sollevamento e granitizzazione. Segue la demolizione della catena, la sedimentazione arenacea (Beacon) dei prodotti dello smantellamento e infine la frammentazione dell'Antartide (e di tutto il Gondwana) con effusione generalizzata di lave (Ferrar).

La superficie orizzontale, che si vede nella parte superiore della fotografia (*in alto*), a oltre 2000 metri di altezza, costituisce il penepiano di Kukry esteso su gran parte dell'Antartide. Esso si impostò sopra una catena sollevatasi nell'Ordoviciano (orogenesi di Ross) e poi smantellata dall'erosione. Si suppone che il penepiano, all'epoca della sua formazione, nel Devoniano, fosse situato poco sopra il livello del mare e che si sia elevato senza deformarsi in epoca recente. Sopra il piano si sviluppano arenarie (Gruppo di Beacon) e quindi effusioni laviche basaltiche (Gruppo di Ferrar). Nelle arenarie del Gruppo di Beacon si riconosce una stratificazione incrociata di origine alluvionale (*in basso*). Questi sedimenti, che formano un ripiano tabulare esteso su gran parte dell'Antartide, sono i prodotti dello smantellamento, da parte dei corsi d'acqua, di una catena montuosa. Vi si trovano resti vegetali: il monte ripreso nella fotografia è stato denominato appunto Timber Peak.

missioni aventi come punto di partenza logistico e organizzativo le basi di Scott e McMurdo Sound, situate 300 chilometri a sud di Baia Terra Nova; le prime missioni utilizzavano ancora i cani da slitta, poi si generalizzò l'uso di mezzi meccanici, soprattutto motoslitte ed elicotteri. La ricerca geologica condotta in questi ultimi anni dagli scienziati italiani è avvenuta in stretta collaborazione con i geologi neozelandesi e con un gruppo tedesco (GANOVEX) che ha costruito una base in prossimità di quella italiana denominandola emblematicamente Gondwana Station.

Il Gondwana

La storia geologica dell'Antartide si inquadra nel contesto della sua appartenenza al supercontinente Gondwana. Sebbene l'idea della deriva dei continenti fosse apparsa in precedenza, fu solo alla fine dell'Ottocento che i geologi ipotizzarono l'esistenza, in tempi geologici remoti, di un grande continente nell'emisfero australe. Questa massa continentale ancestrale, ora chiamata Gondwana, comprendeva gli attuali Sud America, Africa, Arabia, Madagascar, India, Sri Lanka, Antartide e Australia. Data la sua posizione centrale in questo supercontinente, l'Antartide ha avuto un ruolo di primaria importanza nella storia evolutiva del Gondwana. Ciò malgrado, solo recentemente l'esplorazione geologica dell'Antartide ha compiuto passi decisivi per inquadrare questo importante frammento nella cornice già delineata con lo studio degli altri continenti.

Nel 1937 il grande geologo sudafricano Alexander du Toit pubblicò il libro *Our Wandering Continents*, sintesi di fondamentale importanza sulla deriva dei continenti. Malgrado la scarsità di dati allora disponibili sull'Antartide, ne rivendica il ruolo di chiave attorno alla quale possono essere accostati con sorprendente corrispondenza gli altri tasselli del mosaico del Gondwana.

La ricostruzione del Gondwana effettuata e illustrata da Alexander du Toit è basata su parametri geologici confermati da successive ricerche condotte sul continente antartico. Quando il suo libro fu pubblicato, l'ipotesi dell'esistenza del Gondwana aveva ancora pochi sostenitori e il generale riconoscimento della sua opera è avvenuto solamente dopo la sua morte.

Le prime ricerche condotte nelle zone costiere della Penisola antartica e nella fascia antistante il Mare di Ross hanno permesso di individuare due grandi province geologiche: la più estesa comprende gran parte dell'altopiano che si affaccia sugli Oceani Atlantico e Indiano e, poiché si sviluppa nelle aree orientali, è stata chiamata Antartide orientale; essa comprende uno scudo continentale costituito da un basamento cristallino ricoperto da una sequenza sedimentaria e vulcanica a giacitura orizzontale.

L'altra provincia, sul Pacifico, l'Antartide occidentale, è formata da rocce di norma più giovani, intensamente deformate, con un'attività vulcanica che dura fino ai nostri giorni. Essa è chiamata anche Provincia andina in quanto, attraverso la Penisola antartica, costituisce la prosecuzione delle Ande; di fatto si tratta di un grande arcipelago coperto dai ghiacci, comprendente due isole maggiori, la Penisola antartica e la Terra di Marie Byrd.

Le due province sono separate da una catena montuosa - denominata Catena transantartica - geologicamente collegata all'Antartide occidentale. Essa attraversa tutto il continente, con rilievi che possono superare i 4500 metri e offre le informazioni geologiche più interessanti per la storia del Gondwana.

L'origine del Gondwana

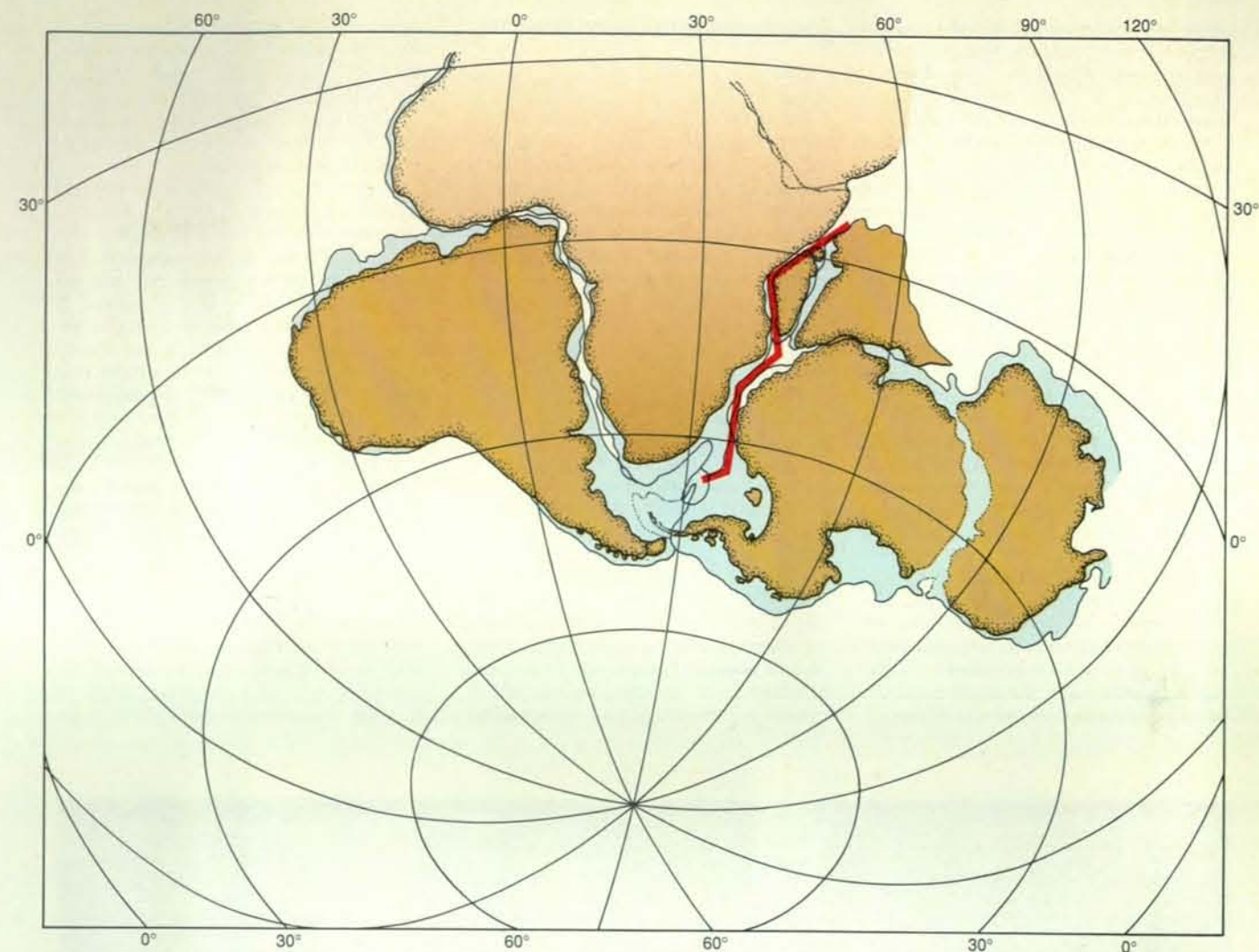
Benché si postuli che il Gondwana esistesse già nel Precambriano (oltre 600 milioni di anni fa), i dati sulla posizione di continenti e oceani nei primi tempi della storia della Terra sono molto scarsi e non sempre attendibili.

In Antartide gran parte dell'altopiano coperto dai ghiacci è rappresentato da uno scudo di rocce cristalline precambriane, lo scudo est-antartico, mentre l'esistenza di uno scudo nell'Antartide occidentale è tuttora in discussione.

Dati recentissimi su questo tema, ricavati dalle osservazioni sul terreno nel settore della Catena transantartica prospiciente la Baia Terra Nova (Wilson Terrane) e valorizzati dallo studio petrografico di campioni raccolti dalle spedizioni italiane, hanno dimostrato che la storia geologica di questo terreno è più complessa di quanto finora ammesso e che esso potrebbe risalire almeno al Proterozoico medio. Sono infatti state scoperte rocce granulitiche analoghe a quelle dello scudo est-antartico, che finora non erano ancora state segnalate nella catena. Esse sono interessanti da un punto di vista mineralogico, poiché registrano condizioni di formazione tipiche della crosta profonda; inoltre testimoniano il coinvolgimento di crosta continentale antica nei processi orogenetici (metamorfismo e piegamento, intrusioni granitiche, sollevamento) del Paleozoico inferiore.

Le prime attendibili ricostruzioni paleogeografiche del Gondwana sono derivate dallo studio delle rocce cambriane (inizio del Paleozoico, oltre 500 milioni di anni fa). Fin dal 1908, la spedizione guidata da Shackleton aveva raccolto campioni di calcari contenenti archeociatidi, fossili cambriani di incerta classificazione, con caratteri affini alle spugne e ai coralli. Recenti ricerche nel settore della Catena transantartica prospiciente la Barriera di Ross hanno dimostrato che tali fossili hanno contribuito alla formazione di una piattaforma prevalentemente calcarea (calcare di Shackleton), attualmente sollevata per oltre 3000 metri, che nel Cambriano si sviluppava ai margini del continente. Queste antiche forme di vita sono state trovate anche in Australia e in altre aree non corrispondenti al Gondwana (Nord America, Russia, Spagna, Sardegna eccetera). Esse testimoniano l'esistenza di climi e ambienti simili a quelli ove attualmente si sviluppano le piattaforme costruite da organismi a guscio calcareo. La ricostruzione del Gondwana nel Cambriano situa infatti l'Antartide in prossimità dell'equatore (*si veda l'illustrazione a pagina 46 in alto*). Se noi riproduciamo su questo antico supercontinente la forma degli attuali continenti, sulla base della loro costituzione geologica, essi compaiono in posizione invertita rispetto all'attuale; non era infatti ancora cominciato il moto di deriva che avrebbe portato l'Antartide ad attraversare il Polo sud nel Siluriano (circa 430 milioni di anni fa) e a raggiungere quindi le alte latitudini meridionali che ha occupato con minori spostamenti dal Carbonifero in poi.

Altri fossili, come trilobiti, echinodermi, molluschi, sono stati rinvenuti nelle rocce sedimentarie cambriane; questi animali vivevano in mari poco profondi, prossimi alla costa, in condizioni simili a quelle delle attuali piattaforme continentali. Sebbene queste associazioni di faune fossili siano analoghe a quelle rinvenute in Australia, Africa e America Meridionale, è impossibile che tali orga-

In questa ricostruzione paleogeografica del Permo-Carbonifero (250 milioni di anni fa) tutti i continenti sono riuniti in un unico grande blocco, il Pangea. Il Gondwana corrisponde al settore australe a sud della Tetide. L'Antartide, vicina al Polo sud, è già caratterizzata da un clima rigido, con espansione di grandi ghiacciai e una vegetazione a felci del genere *Glossopteris*, che si svilupparono anche all'estremità meridionale degli altri continenti che facevano parte del Gondwana. La successione di Gondwana (così chiamata dal nome di una località indiana ove essa è ben rappresentata) di età permo-carbonifera comprende infatti depositi di carbone e sedimenti ricchi di *Glossopteris* alternati a tilliti moreniche.

Così doveva presentarsi il Gondwana nel Giurassico, all'inizio della frammentazione del supercontinente. I singoli blocchi continentali cominciano a delinearsi dapprima con estese fratture da cui fuoriescono grandi quantità di magma di origine sottocrostale, poi con la progressiva espansione dei fondi oceanici corrispondenti. In rosso è indicata la prima frattura che divide il Gondwana in due grandi blocchi. Si noti come nel Giurassico l'Antartide si sia di nuovo allontanata dal Polo sud.

nismi possano avere attraversato oceani paragonabili a quelli che attualmente separano questi continenti: essi dovevano spostarsi in fasce ristrette, cioè in bracci di mare poco profondo che contornavano un'unica massa continentale o si spingevano al suo interno.

Alla suddetta fase di impostazione di una piattaforma continentale ha fatto seguito uno sprofondamento lungo il margine pacifico del Gondwana, corrispondente all'attuale Terra Vittoria, con la formazione di un solco marino stretto e profondo (avanfossa). In questo mare in rapida subsidenza si sarebbero accumulati forti spessori di sedimenti - oggetto di recenti studi da parte delle spedizioni italiane - rappresentati da alternanze di arenarie, argille e calcari provenienti dallo smantellamento dell'area continentale adiacente. È un fenomeno ben conosciuto in altre catene montuose, e precede sempre la nascita della catena stessa. In Antartide si tratta della fase di avanfossa attribuita al Cambriano (anche se la natura detritica della roccia e il susseguente metamorfismo hanno permesso il rinvenimento di rarissimi fossili), cui segue, nell'Ordoviciano, il sollevamento della parte più antica della Catena transantartica (orogenesi di Ross).

L'evoluzione geologica del Gondwana

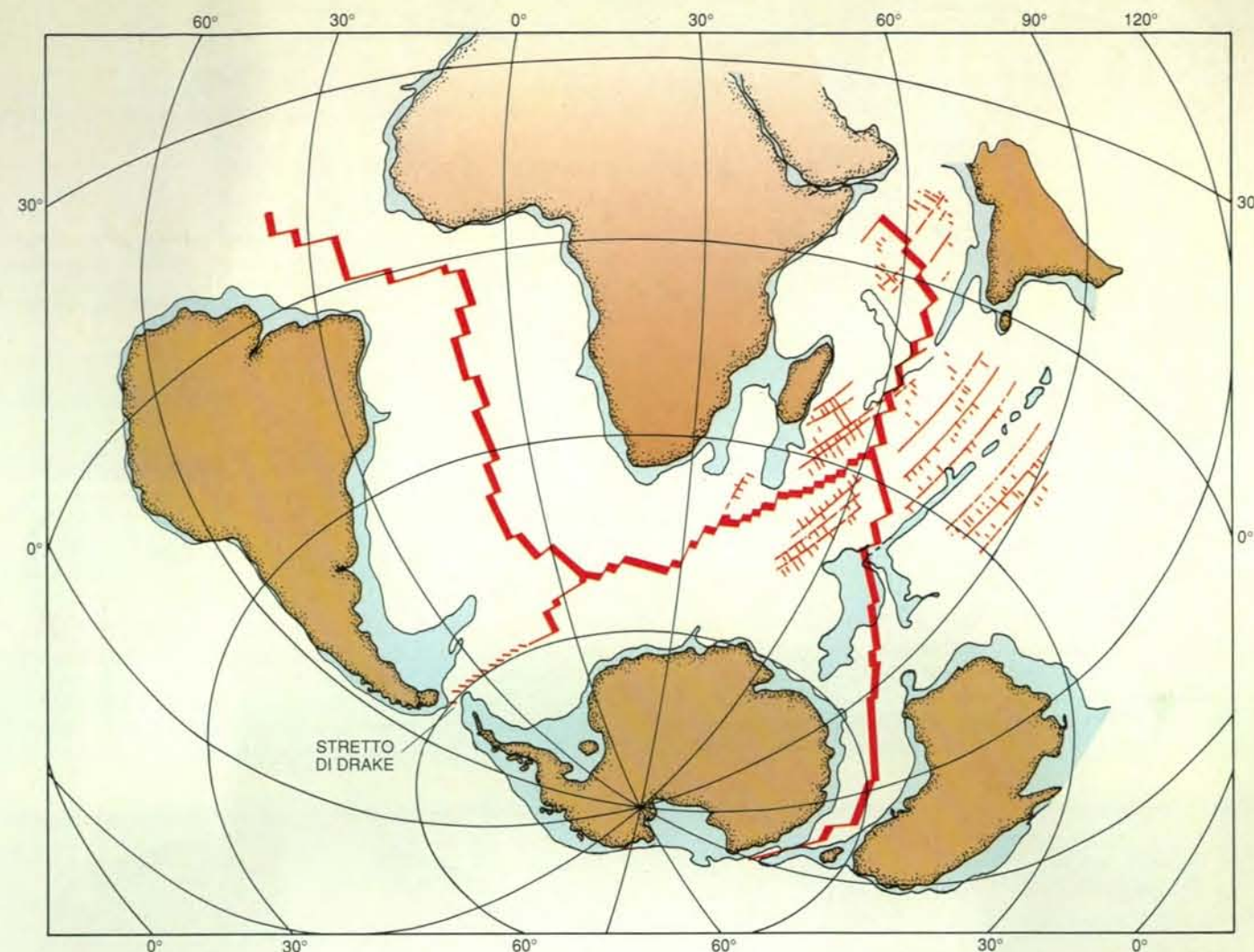
Dal punto di vista geologico l'evoluzione del Gondwana si riassume in tre eventi orogenetici maggiori, documentati da intensa deformazione, attività ignea intrusiva e metamorfismo, che in Antartide si riconoscono lungo la Catena transantartica, sollevatasi appunto in conseguenza di questi eventi, i quali sono ampiamente testimoniati anche nei continenti un tempo adiacenti.

Il primo evento è rappresentato dalla già citata orogenesi di Ross, che può essere ricostruita per quasi tutta l'Antartide in parallelo alla Catena transantartica e che prosegue, da un lato, in Sud Africa, probabilmente con i graniti di Città del Capo, e in Sud America, e, dall'altro, in Australia (orogenesi delameriana).

Nella Terra Vittoria si osservano le testimonianze di questo evento che ha prodotto estesi fenomeni di metamorfismo delle rocce preesistenti e il loro piegamento e sollevamento. Una fitta rete di filoni granitici ed estesi corpi intrusivi attraversa la successione sedimentaria, determinando metamorfismo termico o di contatto (causato da un aumento di temperatura accompagnato da una diminuzione di pressione) riconoscibile sul terreno. Al microscopio si possono osservare minerali caratteristici di queste trasformazioni (andalusite, cordierite, sillimanite). Gli studi radiometrici su questi graniti hanno rivelato un'età di circa 500 milioni di anni che corrisponde all'età dell'orogenesi.

La presenza nella parte nordorientale della Terra Vittoria di rocce sedimentarie cambro-ordoviciane e di corpi granitici datati tra 400 e 360 milioni di anni fa (Devoniano), quindi più giovani di quelli caratteristici dell'orogenesi di Ross, fa sospettare che vi sia registrata un'orogenesi più recente, le cui tracce sembrano presenti anche in alcune aree dell'Antartide occidentale, in particolare nella Terra di Marie Byrd, ma che è soprattutto evidente nell'Australia sudorientale (Orògene Tasman).

Al sollevamento di questa catena è seguito un periodo di erosione fino al suo quasi totale smantellamento. Nella Terra Vittoria i sedimenti che ricoprono l'Orògene di Ross si sono deposti su una superficie orizzontale (penepiano), in netta evidenza morfologica, che oggi si trova normalmente a un'altezza di oltre 2000 metri (*si veda l'illustrazione a pagina 47 in alto*). Questa sovrapposizione di strati orizzontali indeformati, costituiti da arenarie di colore chiaro (Gruppo di Beacon) discordanti su un substrato intensamente ripiegato ed eroso di rocce metamorfiche scure (Gruppo di Wilson) è il fenomeno che più colpisce al primo impatto con la geologia della regione.

Gli espandimenti di lave basaltiche dovuti alla frammentazione della crosta e all'effusione di magma assumono la forma di estese coperture tabulari specie nella zona dell'altopiano glaciale, ove formano vasti *plateau* corrispondenti ai rilievi più alti (Mesa, Monument Nunatak eccetera), con una nomenclatura su base morfologica mutuata dai *plateau* dell'Arizona.

Il Gondwana come doveva apparire nell'Oligocene (30 milioni di anni fa). Durante questo periodo avviene la separazione del blocco continentale antartico dal Sud America e il suo definitivo isolamento con la formazione dello Stretto di Drake.

Le arenarie e le altre formazioni del Gruppo di Beacon si svilupparono a partire dal Devoniano (350 milioni di anni fa) e testimoniano la presenza di una fitta rete fluviale che ha smantellato progressivamente la catena preesistente e ha ridistribuito i sedimenti su buona parte del continente; sono stati rinvenuti pesci fossili devoniani di acqua dolce, che vivevano verosimilmente nei laghi che occupavano le depressioni.

Nel Paleozoico inferiore si verificò il progressivo spostamento di tutto il Gondwana, ancora riunito in un unico supercontinente, verso latitudini polari. Nel Devoniano il Polo sud doveva essere situato nel Sud Africa, mentre l'Antartide, dopo l'attraversamento del polo, si trovava in climi temperati.

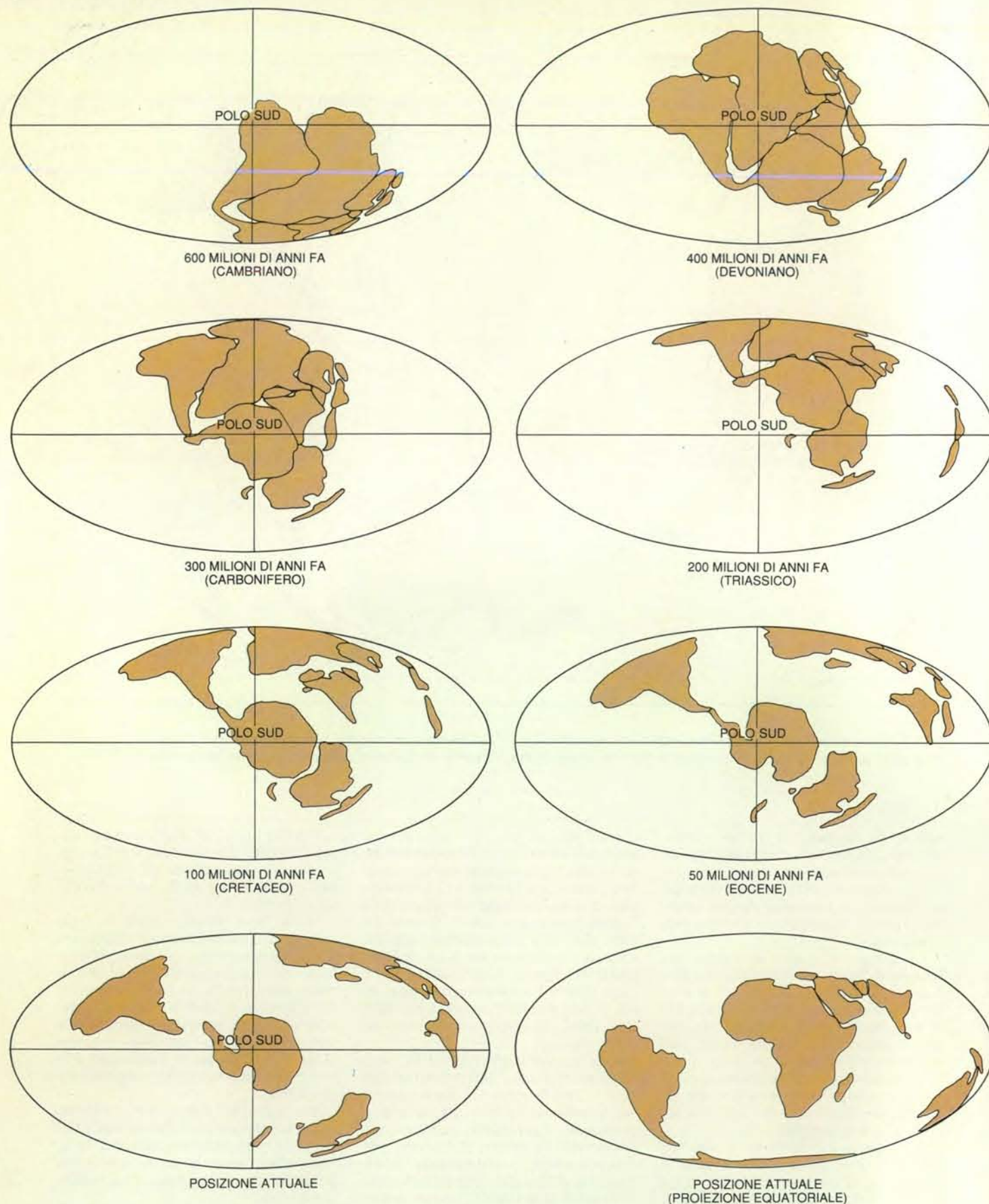
I dati paleoclimatici più sicuri provengono dal Carbonifero (300 milioni di anni fa), con lo sviluppo di una vegetazione caratteristica, le flore a *Glossopteris*. Essa è stata rinvenuta nei Monti della Regina Alessandra che si trovano nel tratto di Catena transantartica sopra l'estremità meridionale del Mare di Ross; queste felci sono infatti indicative di clima freddo e sono state rinvenute anche in tutti gli altri continenti, che facevano parte del Gondwana a partire dal Carbonifero.

Alla fine del Carbonifero e all'inizio del Permiano, circa 280 milioni di anni fa, il raffreddamento del clima culminò nel Gondwana in una glaciazione di estensione continentale, centrata sull'Antartide. La grande glaciazione permo-carbonifera è testimoniata dall'estensione delle tilliti (depositi morenici litificati) e da striature e arrotondamenti dovuti all'azione abrasiva dei ghiacciai. È curiosa la presenza di tilliti carbonifere anche nell'India peninsulare, che nel Gondwana era adiacente all'Antartide.

Nel Permiano, un miglioramento climatico portò alla scomparsa della cappa glaciale, sostituita da ghiacciai montani, e allo sviluppo di nuove flore a *Glossopteris*.

Nella Terra Vittoria, come si è già detto, il Gruppo di Beacon è virtualmente indeformato: esso si presenta sollevato a varie altezze in blocchi di grande estensione, separati da faglie. Viceversa nella Terra di Ellsworth, presso la Penisola antartica, un evento orogenetico permiano, chiamato appunto orogenesi di Ellsworth, coinvolse i sedimenti e le rocce igneo-metamorfiche paleozoiche e precedenti.

Fra la fine del Paleozoico e il Mesozoico si sono verificati importanti cambiamenti nella paleogeografia, con la divisione del Pangea in un continente settentrionale, Laurasia, e uno meridionale, Gondwana.

L'ultima importante orogenesi, quella andina, si sviluppò lungo la fascia del margine pacifico sudorientale, dalla Penisola antartica alla Terra di Ellsworth

La sequenza di proiezioni geografiche polari mostra la posizione delle zolle continentali del Gondwana nei vari periodi geologici mettendo in risalto il ruolo centrale dell'Antartide nell'ambito del supercontinente. (I disegni sono derivati da una elaborazione al calcolatore effettuata da Alessandro Casnedi utilizzando il programma americano «Terra mobilis».)

fino alla Terra di Marie Byrd. Nella Terra di Ellsworth si manifestano i sollevamenti più accentuati (Monte Vinson, 5140 metri). La sua naturale prosecuzione, attraverso la Dorsale pacifico-antartica, è rappresentata dalle Alpi meridionali in Nuova Zelanda.

La frammentazione del Gondwana

I recenti studi geologici e geofisici sia sulle terre emerse sia sui fondi oceanici hanno messo in evidenza le fasi salienti della frammentazione del Gondwana e della separazione dei continenti fino all'assetto attuale.

Più difficile è stabilire esattamente il momento della costituzione dei singoli blocchi continentali, poiché le prime manifestazioni dell'estensione crostale - come attività vulcaniche, formazione di fratture e faglie e trasgressioni marine - possono precedere di molto la definitiva separazione di due blocchi e la creazione di un nuovo fondo oceanico.

In Antartide i primi importanti espandimenti di lave basaltiche provenienti da immense fratture si verificarono nel Giurassico medio (170 milioni di anni fa). Essi documentano l'inizio della separazione del Gondwana in due grandi blocchi: il Gondwana occidentale (Sud America-Africa) e il Gondwana orientale (Antartide-India-Australia). L'attività vulcanica si manifesta attraverso dicchi o filoni che attraversano verticalmente le rocce incassanti; in affioramento questi filoni, più compatti della roccia incassante, formano allineamenti rilevati, facilmente riconoscibili soprattutto con l'osservazione aerea a bassa quota. Spesso essi attraversano filoni granitici preesistenti dai quali si differenziano nettamente per il colore più scuro.

L'evidenza più spettacolare è però rappresentata da immense coltri laviche che sotto forma di spesse colate orizzontali si sovrappongono alle arenarie del Gruppo di Beacon, talvolta con intrusioni del tipo dei filoni a strato. La composizione è basaltica, caratteristica del vulcanismo da magmi di provenienza sottocrostale. Queste coperture basaltiche (Gruppo di Ferrar) si sviluppano regolarmente specie nella zona dell'altopiano glaciale (*si veda l'illustrazione a pagina 50*).

In epoca di poco successiva (tardo Giurassico-Cretaceo) i due blocchi nei quali si era smembrato il Gondwana si frammentarono ulteriormente, dando luogo agli odierni continenti con una forma abbastanza simile a quella attuale. L'Africa si separò dal Sud America con la formazione dell'Atlantico, che iniziò a espandersi a partire dalla sua estremità meridionale.

Secondo le ipotesi più accreditate, basate sulle curve paleomagnetiche, sulle anomalie magnetiche dei fondi oceanici e su dati geologici, il processo di *rifting* e il vulcanismo iniziarono durante il Giurassico, mentre la formazione di fondo oceanico cominciò nel Cretaceo.

In età analoghe, anche se esistono pareri discordi, dovrebbe essersi verificata una prima frammentazione del Gondwana orientale con la separazione dell'India dall'Antartide e l'inizio della rapida espansione dell'Oceano Indiano che ha portato, in tempi più recenti, l'India in collisione con l'Asia. Con la creazione di un ramo laterale della Dorsale dell'Oceano Indiano (Dorsale indiana sudorientale) si sarebbe in seguito (85 milioni di anni fa) separata dall'Antartide anche l'Australia con la Nuova Zelanda, lasciando in collegamento col blocco antartico l'esteso *plateau* delle Isole Kerguelen.

L'isolamento dell'Antartide

La frammentazione del Gondwana nei suoi due blocchi orientale e occidentale, durante il Giurassico, ha però lasciato una connessione geologica fra l'estremità meridionale del Sud America e la Penisola antartica, che presentano una struttura e una successione di terreni molto simili; alcuni autori ritengono che all'inizio del Terziario una stretta lingua di terra unisse i due continenti. Il definitivo distacco, con l'isolamento dell'Antartide e l'apertura dello Stretto di Drake, avvenne probabilmente solo nell'Oligocene, vale a dire 30 milioni di anni fa. Queste interpretazioni sono avvalorate dai dati emersi dai sondaggi del Deep Sea Drilling Project che hanno perforato e campionato i fondi oceanici del bacino del Pacifico sudorientale e i suoi margini continentali. Il substrato basaltico raggiunto mostra una crosta oceanica non più antica del Cretaceo superiore, ma si suppone che tale bacino esistesse fin dal Paleozoico. Sono evidenti discordanze di età oligocenica, con erosioni dovute all'inizio della circolazione della Corrente circumpolare antartica.

Con la definitiva separazione dell'Antartide si attivò infatti attraverso lo Stretto di Drake una circolazione di importante impatto climatico che contribuì all'isolamento anche biologico del continente. Pertanto, sebbene si supponga che l'Antartide si sia trovata in posizione polare o circumpolare fin dall'inizio del Mesozoico, è solo da circa 25 milioni di anni, e cioè dalla fine dell'Oligocene, che si è sviluppata l'immensa glaciazione che ne ha quasi completamente coperto la superficie.

La struttura del continente è stata recentemente modificata da importanti eventi come l'apertura del Mare di Ross, di cui esiste la testimonianza più diretta sotto forma di una grande frattura che borda questo mare nel lato prospiciente la Terra Vittoria; la prova più evidente di questa frattura è l'allineamento di vulcani che da Cape Adare, attraverso il Monte Melbourne nella Baia Terra Nova, arriva fino al Monte Erebus, presso la base di McMurdo.

L'antico Gondwana è rappresentato dalla parte orientale dell'Antartide, la cui struttura crostale, tipicamente continentale, con uno spessore medio di circa 40 chilometri, ha origine dai primordi della storia della Terra. L'Antartide orientale è circondata da dorsali oceaniche, ad attivo vulcanismo, che la separano dagli altri frammenti dell'antico supercontinente Gondwana ed è bordata a occidente dalla Catena transantartica che testimonia le fasi più antiche della evoluzione strutturale del continente. L'Antartide occidentale invece è formata da catene mesozoiche e terziarie sollevatesi in seguito alla collisione della zolla antartica con quella del Pacifico. La recente espansione della crosta dell'Antartide ha causato il suo assottigliamento in corrispondenza delle depressioni dei Mari di Ross e Weddell e profonde fratture lungo le quali si allineano vulcani attivi.

Nella storia del Gondwana, l'Antartide rappresentò il settore centrale del supercontinente fino alla sua frammentazione giurassica. Si suppone che nella prima parte della sua storia geologica fosse situata a latitudini equatoriali; nel Carbonifero (300 milioni di anni fa), l'Antartide era già in posizione prossima al polo, all'estremità meridionale di un'unica grande massa continentale, il Pangea. Dopo la separazione dei due supercontinenti Laurasia e Gondwana, quest'ultimo ha iniziato a frammentarsi oltre 170 milioni di anni fa dando luogo agli attuali continenti australi e all'India, che si è spostata progressivamente verso nord in seguito all'apertura dell'Oceano Indiano. La zolla antartica ha raggiunto, in vari moti di deriva, la sua attuale posizione polare circa 30 milioni di anni fa, staccandosi definitivamente dagli altri continenti. Questo isolamento geografico e climatico ha contribuito alla formazione della calotta glaciale che ricopre il 98 per cento della sua superficie.

BIBLIOGRAFIA

DU TOIT ALEXANDER, *Our Wandering Continents; An Hypothesis of Continental Drifting*, Oliver and Boyd, Edimburgo, 1937.

CRADDOCK C., *Antarctica and Gondwanaland* in *Antarctic Geoscience*, University of Wisconsin Press, Madison, 1982.

AUTORI VARI, *Geosciences in Victoria Land, Antarctica*, a cura di Carlo Alberto Ricci in «Memorie della Società Geologica Italiana», XXXIII, Siena, 1987.

AUTORI VARI, *Gondwana Six: Structure, Tectonic and Geophysics*, a cura di G. D. McKenzie, American Geophysical Union, Washington, 1987.

AUTORI VARI, *Earth Science Investigations in Antarctica*, a cura di Carlo Alberto Ricci in «Memorie della Società Geologica Italiana», XLVI, Siena, 1991.

La macchia cieca della retina

Lo studio di questa struttura consente di comprendere come il cervello umano elabora le immagini riuscendo a compensare le lacune percettive

di Vilayanur S. Ramachandran

Nel XVII secolo, lo scienziato francese Edmé Mariotte, dissezionando un occhio umano, scoprì la papilla ottica, ossia la zona in cui il nervo ottico entra nel bulbo oculare espandendosi a formare la retina. Si accorse anche che, diversamente dal resto della retina, quella parte era insensibile alla luce. Sulla base delle proprie conoscenze di ottica e di anatomia, Mariotte giunse allora alla conclusione che ogni occhio, in una piccola porzione del campo visivo, presenta una macchia cieca.

Il lettore può confermare senza difficoltà la scoperta di Mariotte osservando il disco su fondo colorato riprodotto in questa pagina. Chiudendo l'occhio destro e tenendo la pagina a circa 30 centimetri dal viso, concentri lo sguardo sul quadrato mentre avvicina lentamente il foglio all'occhio sinistro. A una certa distanza critica, il disco viene a cadere nella macchia cieca e sparisce completamente. In realtà, viene percepito come se fosse stato ricoperto (riempito) dal colore del fondo. Questo processo visivo è noto come riempimento (*filling in*).

Di recente, si è cominciato a comprendere che il fenomeno sopra descritto è una manifestazione di un meccanismo percettivo più generale, l'interpolazione di superficie. Per esempio, quando un individuo osserva un tavolo, è probabile che il suo sistema visivo estragga alcune informazioni sui bordi e crei, di quel tavolo, una rappresentazione mentale simile allo schizzo che comparirebbe in una vignetta. Il sistema visivo potrebbe quindi ricorrere all'interpolazione di superficie per completare con il colore e la venatura del legno l'immagine del tavolo, semplificando così il processo di percezione dei dettagli della superficie.

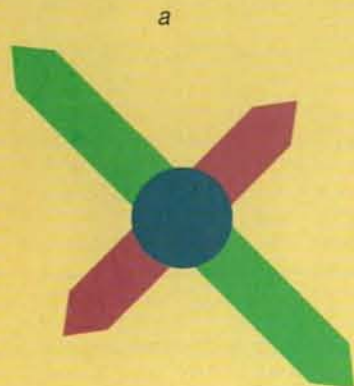
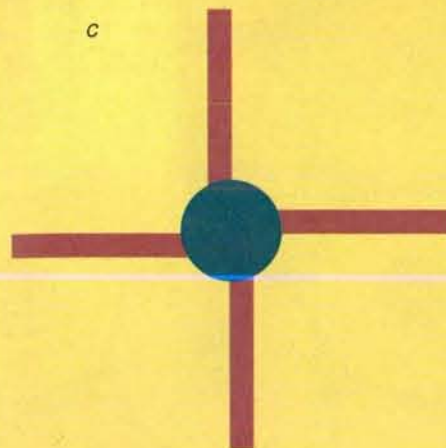
Per comprendere come il cervello interpoli l'informazione visiva, i miei collaboratori e io, all'Università della California a San Diego, abbiamo effettuato molti esperimenti per analizzare la percezione nella macchia cieca della retina. Le illustrazioni che corredano questo articolo offrono al lettore l'opportunità di provarne alcuni. Abbiamo anche indagato su due tipi di macchie cieche «artificiali», dapprima mettendo a punto una tecnica che era in grado di indurre, senza alcun rischio, macchie cieche temporanee e poi esaminando parecchi individui che soffrivano di scotomi (cioè presentavano zone cieche causate da un danno a una piccola porzione del cervello). Questa ricerca ci ha permesso di scoprire molte caratteristiche del fenomeno di riempimento e abbiamo così cominciato a comprendere come esista una correlazione tra questo processo e altri processi visivi, quali il riconoscimento dei bordi e la percezione del moto.

Da tempo è noto che, se un individuo guarda un oggetto, nelle aree visive del cervello si forma una rappresentazione percettiva di quell'oggetto. Viceversa, il cervello ricorre a uno stratagemma per non perdere contatto con gli oggetti che si trovano al di fuori del campo visivo (per esempio, gli oggetti situati dietro la testa). Per essi il cervello crea quella che, in senso lato, potrebbe essere chiamata una rappresentazione concettuale, cioè qualcosa di analogo a una deduzione logica. La distinzione non è semplicemente semantica. Le rappresentazioni percettive e concettuali vengono generate probabilmente in aree corticali distinte e possono essere elaborate in modi molto diversi.

Quanto è ricca la rappresentazione percettiva nella regione corrispondente alla macchia cieca? Per rispondere a questa domanda, abbiamo chiesto ad alcuni volontari di esaminare una serie di immagini semplici. Siamo partiti da un'immagine in cui una banda è interrotta da un disco (*si veda l'illustrazione in questa pagina*). Quando uno dei volontari ha orientato la propria macchia cieca in modo che ricoprisse il disco, la banda gli è apparsa continua sia come forma sia come colore. Sono state effettuate osservazioni analoghe quando è stato mostrato un segmento verticale rosso in alto e verde in basso, ricoperto da un disco nel punto in cui i due colori si incontravano. I volontari hanno riferito che, quando il disco andava a cadere nella macchia cieca, il segmento appariva continuo anche se, paradossalmente, essi non potevano in realtà vedere il limite tra rosso e verde. Il paradosso nasce presumibilmente dal fatto che una parte del sistema visivo segnala che il segmento è continuo, mentre un'altra parte non è in grado di discernere il limite tra il colore rosso e il colore verde.

Per verificare se il sistema visivo è in grado di completare anche immagini più complicate, ho escogitato insieme a mia moglie, Diane Rogers-Ramachandran, parecchie figure, tra cui lo schema di una ruota di bicicletta (*si veda l'illustrazione in basso nella pagina successiva*). Quando la macchia cieca è orientata sul mozzo della ruota, sembra che i raggi convergano in un unico punto al centro della macchia.

Un effetto analogo si può osservare se il centro di una grossa X cade nella macchia cieca. La X appare completa. Ma, fatto di notevole rilievo, se uno dei segmenti che la compongono è più lungo dell'altro, solo esso appare completo (*si veda l'illustrazione in alto nella pagina successiva*). Questo risultato indica che il processo di riempimento potrebbe essere influenzato da stimoli visivi che provengono da punti distanti dalla

Il disco (*nella pagina a fronte*) tende a sparire se viene posizionato all'interno della macchia cieca della retina. Si chiuda l'occhio destro e si tenga la pagina a una distanza di circa 30 centimetri. Si fissi lo sguardo sul quadrato e, contemporaneamente, si avvicini a poco a poco la pagina all'occhio aperto. A un certo punto il disco scomparirà e verrà ricoperto dal colore del fondo. Anche i dischi rappresentati qui sotto spariscono e vengono ricoperti dal colore del fondo se cadono all'interno della macchia cieca, mentre la striscia a sinistra apparirà completa. Lo stesso si può dire per quella al centro, anche se il colore della parte in cui si è avuto il riempimento non dovrebbe essere riconoscibile. Il tratto di striscia dell'ultima figura a destra non dovrebbe sporgere nella macchia cieca.

Il processo di riempimento, associato alla macchia cieca della retina, è una funzione visiva complessa. Se la macchia cieca cade al centro di una croce (*a*), solo il segmento più lungo di questa apparirà completo. Nel caso di un insieme di dischi (*b*), il disco che va a coincidere con la macchia cieca semplicemente svanisce nel fondo. In *c* i segmenti verticali e orizzontali non sono allineati, ma quando il disco viene a trovarsi nella macchia cieca, la striscia verticale sembra continua e rettilinea e quella orizzontale appare sfalsata. Questo effetto è stato notato per la prima volta da Jerome Lettvin e collaboratori della Rutgers University. Anche il movimento influisce sulla percezione associata con la macchia cieca. Se si passa di continuo dall'illustrazione *d* a quella in alto a pagina 61 e viceversa, si ha l'impressione che la striscia si muova in diagonale.

macchia cieca. Siamo stati anche in grado di dimostrare che il riempimento non è un processo «cognitivo», cioè non si basa sulle attese di quello a cui i vari oggetti dovrebbero assomigliare. Per esempio, abbiamo generato un'immagine consistente in una colonna verticale di dischi, uno dei quali va a cadere nella macchia cieca (*si veda b nell'illustrazione di questa pagina*). Tutti i soggetti hanno visto svanire quel disco e non hanno avuto l'«allucinazione» del disco mancante per conservare l'immagine completa.

Una delle osservazioni più interessanti si è avuta quando abbiamo compiuto esperimenti con contorni illusori. Il sistema visivo produce questi contorni quando l'occhio registra certi insiemi di forme incomplete o di bande discontinue. Per esempio, abbiamo generato un'immagine consistente in parecchie bande orizzontali su fondo colorato, ognuna delle quali si interrompe a un terzo. Il sistema visivo non interpreta l'immagine come un insieme di bande discontinue, ma produce contorni illusori che definiscono una striscia verticale.

Quando abbiamo chiesto ai volontari di orientare la propria macchia cieca su una interruzione lungo una delle bande orizzontali, i più hanno riferito che la striscia verticale appariva completa nella zona della macchia cieca, mentre quella orizzontale rimaneva discontinua. Successivamente, presentando la stessa immagine in cui erano state tolte tutte le bande tranne tre, quella che cadeva nella macchia cieca è apparsa completa. Pertanto sembra che il processo di riempimento dipenda dal grado di definizione dei contorni illusori.

La figura della ruota appare completa quando il disco blu cade nella macchia cieca.

La neurofisiologia può aiutare a comprendere il meccanismo alla base del riempimento. Per prima cosa, le cellule della retina sensibili alla luce traducono l'intensità luminosa e il colore in impulsi elettrici che sono inviati alla corteccia visiva primaria tramite il nervo ottico. In quest'area i neuroni smistano l'informazione visiva e l'inviano a parecchie altre aree, ciascuna delle quali sembra essere specializzata per un tipo di elaborazione visiva, riguardante per esempio il colore, il movimento e forse la forma.

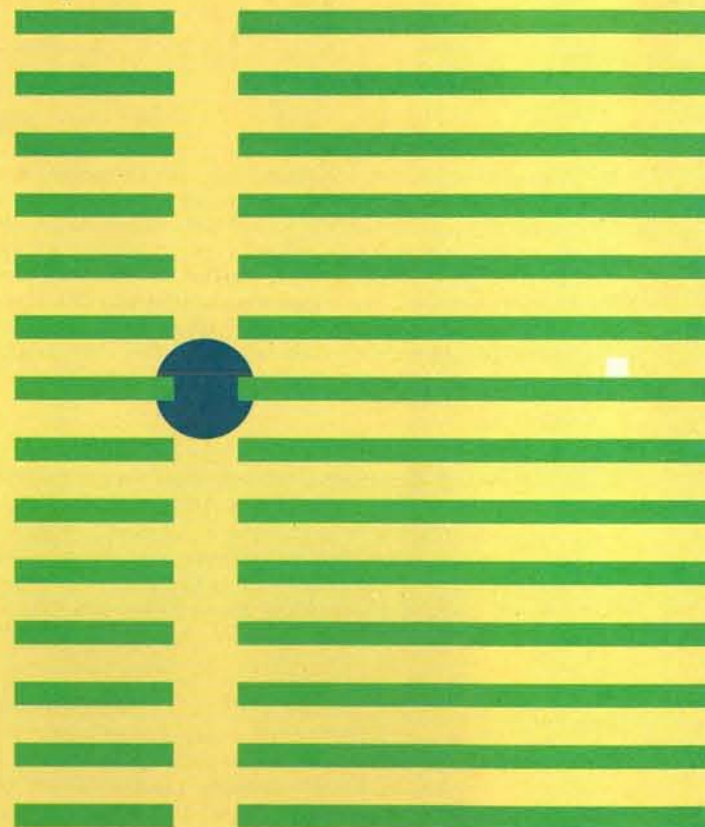
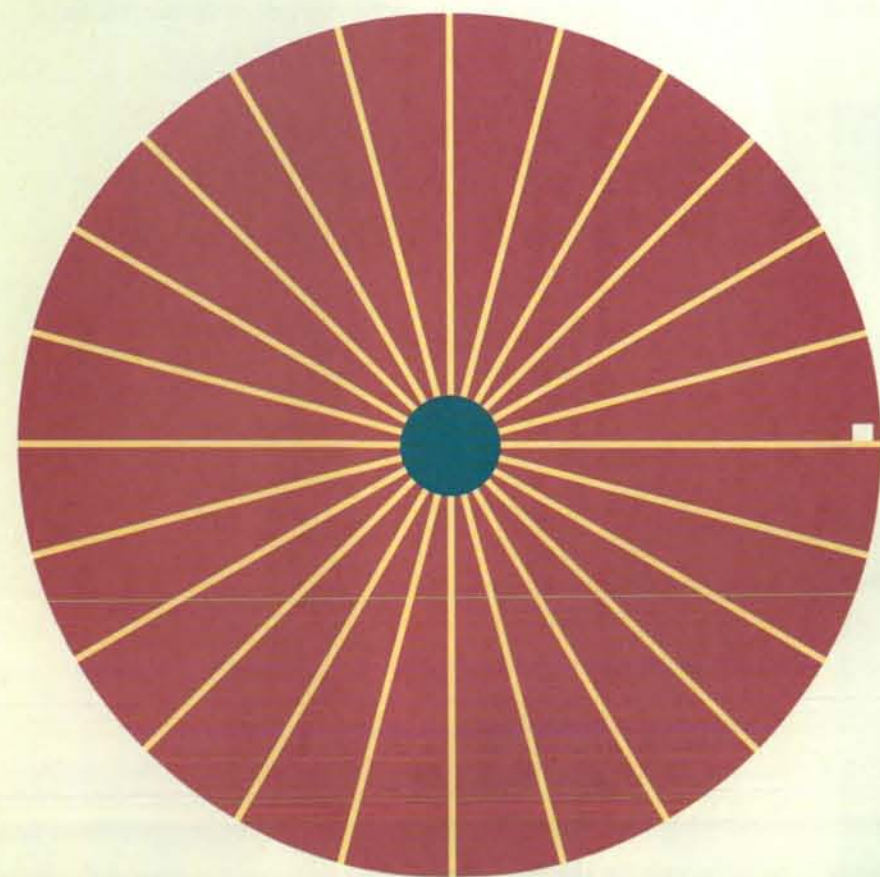
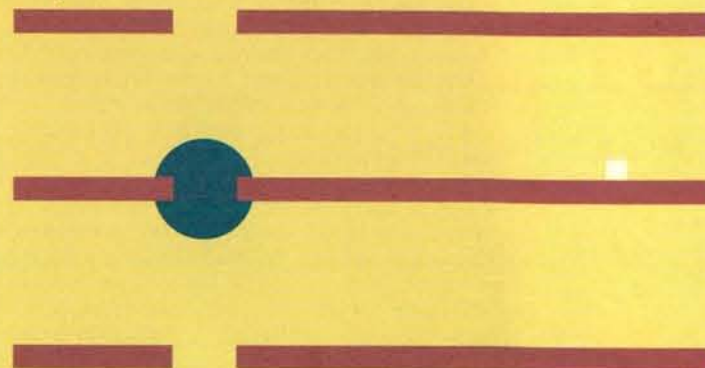
A partire dal 1987 i miei collaboratori e io abbiamo cercato di determinare se il riempimento abbia luogo prima, dopo o durante altri tipi di elaborazione visiva. Siamo partiti da un tipo piuttosto semplice di elaborazione che provoca l'effetto percettivo di far «risaltare» (*pop-out*) qualche aspetto della visione. Il sistema visivo tende a dirigere l'attenzione verso determinate caratteristiche elementari dell'ambiente visivo, che spiccano rispetto ad altre forme prive d'interesse e diverse per colore od orientazione. Secondo alcuni psicologi, l'effetto sopra descritto avviene in realtà solo nei confronti delle caratteristiche che, nel processo di elaborazione visiva, vengono estratte in una fase relativamente precoce.

Ma se si fa coincidere la macchia cieca con l'interruzione e poi si passa da una all'altra delle due pagine, la striscia pare muoversi solo in senso verticale. In *e* i contorni soggettivi danno l'illusione di una striscia verticale posta tra bande orizzontali. La striscia verticale appare continua quando il disco cade nella macchia cieca. Se essa non è ben definita (*f*), è la banda orizzontale che cade nella macchia cieca a essere completata.

I miei collaboratori e io ci siamo chiesti se il riempimento abbia luogo prima o dopo il processo che fa risaltare determinati aspetti del quadro. Per indagare su questo punto abbiamo generato un'immagine costituita da parecchi anelli e abbiamo chiesto ad alcuni volontari di far coincidere il centro di uno di questi anelli con la loro macchia cieca. Essi hanno fatto un'osservazione interessante: l'anello che circondava la macchia cieca veniva riempito e trasformato in un disco omogeneo; hanno anche notato che «risaltava» in mezzo agli altri.

Questa osservazione ha due implicazioni importanti. In primo luogo, il processo di riempimento deve avvenire precocemente durante l'elaborazione visiva, dato che precede quello che fa risaltare determinati aspetti del quadro. Inoltre esso deve comportare la generazione di una rappresentazione percettiva, senza la quale il secondo processo non avverrebbe.

William Aiken e io abbiamo quindi esaminato la relazione tra riempimento e percezione del moto, che si instaura nei primissimi stadi del processo di elaborazione visiva. Ci siamo basati su un'illusione ben nota, quella del moto apparente. Se, per esempio, due linee parallele identiche vengono mostrate rapidamente l'una dopo l'altra, il cervello tende a percepire un'unica linea in movimento. Il moto apparente è, in effetti, l'illusione che ci permette di percepire un movimento continuo quando, su uno schermo cinematografico, viene proiettata una serie di immagini.

Per i nostri esperimenti abbiamo ideato due immagini, ciascuna contenente una banda interrotta (*si vedano le illustrazioni nell'angolo in alto a destra di questa pagina e di pagina 61*). Abbiamo posizionato le interruzioni in modo che, quando le immagini venivano presentate in rapida successione, l'interruzione sembrava spostarsi in diagonale. Per cominciare, abbiamo presentato a un soggetto la prima immagine e gli abbiamo chiesto di orientare la macchia cieca sull'interruzione. Quando ci ha detto che la banda appariva completa, abbiamo

In mezzo a un gruppo di segmenti verticali un segmento diagonale si distingue nettamente, mentre una L in un gruppo di T è più difficile da individuare. Il segmento diagonale emerge per il processo percettivo di *pop-out* (che fa risaltare un elemento del quadro). Per lo stesso fenomeno, se il centro di uno degli anelli riprodotti in *b* è posizionato nella macchia cieca, l'anello viene riempito e risalta rispetto agli altri. Se si fissa lo sguardo per parecchi secondi sul quadratino dell'illustrazione *c*, i colori sembrano riempire il disco da entrambi i lati.

tolto quella prima immagine e abbiamo proiettato la seconda immagine spostata. Ci saremmo aspettati che, se il riempimento avesse avuto luogo dopo la percezione del moto, il soggetto avrebbe continuato a vedere il movimento in diagonale. In realtà, egli ha riferito che le bande sembravano muoversi in verticale. Questi e altri esperimenti ci hanno portato alla conclusione che il riempimento è un processo che precede la percezione del moto.

È difficile talvolta effettuare osservazioni accurate della macchia cieca della retina perché essa si trova sempre alla periferia del campo visivo dell'occhio. Con Sir Richard Gregory dell'Università di Bristol abbiamo tentato di risolvere questo problema e abbiamo trovato un metodo innocuo per indurre una macchia cieca artificiale vicino al centro del campo visivo dell'occhio.

Il lettore può ripetere i nostri esperimenti a casa propria servendosi di un televisore. Scelga un canale non sintonizzato, in modo che sullo schermo compaia l'effetto «neve», cioè uno sfavillio di punti. Incolli quindi al centro dello schermo una piccolissima etichetta circolare. A circa otto centimetri da questa fissi un quadrato di carta di colore grigio, di un centimetro di lato e con una luminanza grosso modo corrispondente al grigio dell'effetto neve. Per produrre lo stesso effetto visivo, Gregory e io abbiamo utilizzato un calcolatore per generare le forme e lo sfavillio di punti.

Se si fissa l'etichetta per circa 10 secondi, ci si accorge che il quadrato scompare completamente e viene «sostituito» dallo sfavillio di punti. Secondo noi, questo processo di riempimento è analogo a quello associato alla macchia cieca reale e può basarsi su meccanismi nervosi simili. La scomparsa del quadrato è probabilmente dovuta a una iperstimolazione con conseguente affaticamento dei neuroni che segnalano la presenza del quadrato. Normalmente ciò non avviene perché l'occhio si muove tutt'attorno, evitando una stimolazione eccessiva.

Questi risultati concordano con esperimenti di fisiologia eseguiti di recente da Charles Gilbert e Torsten Wiesel della Rockefeller University. Studiando i primati, i due ricercatori hanno effettuato una scoperta notevole: le cellule della retina che circondano una zona lesa, insensibile alla luce, possono cominciare molto rapidamente a influenzare la corteccia visiva primaria nella regione associata a quella stessa area. Questa osservazione (e altre effettuate da Ricardo Gatas dei National Institutes of Health e da John Kaas della Vanderbilt University) potrebbe spiegare il processo di riempimento che abbiamo osservato.

Di recente siamo riusciti a realizzare un'interessante variazione dell'esperimento di sfavillio di punti. Quando un volontario ci ha comunicato che il quadrato era stato riempito dai punti, abbiamo chiesto al calcolatore di far diventare uniformemente grigio lo schermo. Con nostra sorpresa, tutti i volontari sottoposti all'esperimento hanno riferito di vedere, nella regione in cui il quadrato grigio originario era stato riempito, una chiazza quadrata di punti luminosi che rimaneva visibile anche per 10 secondi.

L'osservazione fa pensare che, in realtà, un insieme di neuroni generi una rappresentazione della regione riempita da uno sfavillio di punti. Inoltre è evidente che tale rappresentazione può persistere anche dopo che i punti circostanti sono scomparsi.

Gli effetti del riempimento si possono osservare anche in immagini statiche. Per esempio, i confini tra zone di colore diverso tendono a svanire quando il soggetto fissa una particolare immagine. (Il fatto è stato studiato in modo approfondito da A. L. Yarbus dell'Accademia delle scienze di Mosca e da Thomas Piantanida della SRI International.) Mio figlio, Chandramani Ramachandran, e io abbiamo effettuato un esperimento con un'immagine in cui un disco grigio sta a cavallo di una linea di demarcazione verticale tra due colori, grigio e blu, aventi uguale luminanza. Abbiamo chiesto ai volontari di guardare il disco per parecchi secondi. Essi hanno riferito che risultava riempito da entrambi i colori, ma che non riuscivano a distinguere una linea di demarcazione tra questi. Al contrario, vedevano una nebulosità diffusa.

Successivamente, abbiamo alterato un poco l'immagine aggiungendo parecchi segmenti orizzontali neri. Era nostra intenzione creare un contorno illusorio che coincidesse con la linea di demarcazione tra i colori. Dopo che i soggetti hanno osservato l'immagine per alcuni secondi, i colori hanno riempito il disco da entrambi i lati, ma, diversamente dall'esperimento precedente, hanno formato una linea di demarcazione netta tra i colori, lungo il contorno illusorio. Questo risultato è sorprendente in quanto la maggior parte dei ricercatori pensa che il riempimento dipenda da fattori semplici come la presenza di bordi definiti da cambiamenti nella luminanza.

Lo scorso anno abbiamo avviato ricerche su uno dei tipi più interessanti di macchia cieca. Se, a seguito di un incidente, di una malattia o di un intervento chirurgico, in un individuo rimane lesa una minuscola zona della corteccia visiva, si dice che egli presenta uno scotoma, ossia è cieco in un piccolo settore del campo visivo.

Spesso il soggetto è completamente inconsapevole di questo fatto. Se, per esempio, osserva un motivo su una tappezzeria, non percepisce lo scotoma come cambiamento di colore o come interruzione nel disegno; anzi, il disegno gli appare uniforme. Eppure, se un disco viene incollato alla tappezzeria nell'area corrispondente allo scotoma, l'individuo non riesce a notare il disco e continua a vedere soltanto la tappezzeria.

I miei collaboratori e io abbiamo esaminato due pazienti con scotomi (dovuti a lesioni delle aree visive del cervello) in prossimità del centro del campo visivo. Abbiamo innanzitutto mostrato ai pazienti un grosso cerchio, che si sovrapponeva parzialmente allo scotoma. Essi hanno riferito che, dopo circa otto secondi, la parte oscurata del cerchio risaltava, completando la figura. Questo effetto contrasta nettamente con esperimenti analoghi sulla macchia cieca naturale: quando una parte del cerchio cade in essa, il riempimento non ha luogo.

Perché questa differenza? La maggior parte dei neuroni della corteccia visiva riceve impulsi dalle cellule retiniche di entrambi gli occhi, ma la corteccia contiene anche un gruppo di neuroni che corrisponde alla macchia cieca dell'occhio sinistro e che riceve segnali soltanto dall'occhio destro (e viceversa). Pertanto, se un cerchio va a situarsi in parte nella macchia cieca, per esempio dell'occhio sinistro, si ha di solito una compensazione da parte dell'occhio destro che segnala la presenza della parte di cerchio invisibile all'occhio sinistro. Il sistema visivo percepisce così un cerchio completo. Ma, se l'occhio destro è chiuso, non segnala la presenza della parte oscurata e il sistema visivo suppone che manchi un pezzo di cerchio.

D'altro canto, se la corteccia visiva viene distrutta in qualche sua parte, non riesce più a elaborare i segnali che provengono sia dall'uno sia dall'altro occhio e sembra allora che il sistema visivo si adatti ricorrendo al processo di riempimento.

Abbiamo poi generato un'immagine di una striscia verticale interrotta da un disco. I pazienti hanno riferito che, quando lo scotoma era orientato sul disco, la striscia veniva completata, anche se, stranamente, il processo richiedeva circa cinque secondi. Abbiamo quindi spostato orizzontalmente la sezione superiore della striscia in modo che le parti ai due lati del disco risultassero sfalsate. I pazienti hanno riferito che all'inizio le bande apparivano disallineate, ma ben presto cominciavano a muoversi orizzontalmente l'una verso l'altra fino a diventare allineate e a connettersi attraverso lo scotoma. Hanno anche riferito che la sensazione di movimento era molto viva.

Possiamo solo avanzare ipotesi su questo fenomeno di riallineamento. Una delle aree visive superiori può disporre di qualche indizio che segnala che i segmenti situati ai due lati dello scotoma fanno parte della stessa linea. Quando essa non riceve segnali contrastanti dalla corteccia visiva primaria (in quanto danneggiata), il sistema visivo può finire per interpretare l'immagine come una striscia verticale unica.

Infine, Kerrie A. Maddock, Daniel Plummer e io abbiamo chiesto a uno dei pazienti di guardare uno schermo televisivo in cui compariva uno sfavillio di punti rossi. All'inizio il paziente ha riferito che lo scotoma veniva riempito solo dal colore rosso, mentre dopo circa otto secondi comparivano anche i punti. Questo esperimento indica che l'area visiva responsabile del riempimento per il colore possa essere diversa dall'area responsabile del riempimento per il movimento.

In avvenire speriamo di riuscire a chiarire affinità e differenze tra scotomi e macchie cieche. Forse riusciremo a scoprire sia la correlazione tra processo di riempimento e interpolazione di superficie sia le aree sedi di tali processi.

I lettori potrebbero contribuire alla ricerca continuando gli esperimenti con la macchia cieca. Con un po' di pratica dovrebbero riuscire a orientarla in modo da far scomparire qualunque piccolo oggetto. Si dice che Carlo II di Inghilterra si dilettasse in questo modo, decapitando visivamente le dame di corte. Spero che i lettori scelgano un altro modo di divertirsi.

BIBLIOGRAFIA

MARR DAVID, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman and Company, 1982.

GREGORY RICHARD L., *Occhio e cervello*, Raffaello Cortina, Milano, 1991.

RAMACHANDRAN V. S. e GREGORY R. L., *Perceptual Filling in of Artificially Induced Scotomas in Human Vision* in «Nature», 350, n. 6320, 25 aprile 1991.

RAMACHANDRAN V. S., *Filling in Gaps in Perception* in «Current Directions in Psychological Science» (in stampa).

Ottica binaria

Il connubio fra ottica e microelettronica ha già prodotto lenti aventi lo spessore di un quarantesimo di questa pagina e schiere di 10 000 telescopi, ciascuno non più grande del diametro di un capello umano

di Wilfrid B. Veldkamp e Thomas J. McHugh

Entrare in un laboratorio di ottica di precisione fa pensare di essere tornati a un'altra epoca: l'ambiente è pieno di macchine pesanti, cinghie, pulegge, scatole di latta, pennelli, lavelli imbrattati, fornelli improvvisati. Artigiani esperti usano mole e strumenti rivestiti con strati di materiale abrasivo via via più fine per realizzare lenti per telescopi, strumentazione ottica per l'elaborazione di segnali e altri sistemi. Sebbene tali metodi possano fornire lenti di eccellente qualità, di forma precisa fino a una piccola frazione di lunghezza d'onda, la produzione è lenta e richiede la massima cura. Inoltre la molatura e la levigatura tradizionali consentono la realizzazione soltanto di forme sferiche, che non focalizzano la luce con la massima efficienza.

L'ottica binaria (così detta per il suo stretto legame con la costruzione dei circuiti digitali) è una nuova tecnologia che elimina molte fasi della produzione tradizionale delle lenti. In contrasto con il metodo convenzionale, si incidono direttamente le forme volute sulla superficie di un materiale ottico, con l'impiego di tecniche sviluppate in origine per la realizzazione di circuiti integrati. L'ottica binaria produce lenti di alta qualità e consente la riproduzione rapida e a basso costo di elementi ottici le cui proprietà di messa a fuoco possono essere adattate in modi un tempo impossibili.

Grazie a questa tecnologia, i progettisti non sono più vincolati dai limiti dell'ottica convenzionale. Essi possono creare virtualmente qualunque sistema o componente, entro i limiti dell'ingegnosità umana, purché questo possa essere realizzato con l'attuale tecnologia microcircuitale. Inoltre l'ottica binaria potrà progredire sfruttando il progressivo miglioramento dei metodi di fabbricazione dei microcircuiti. Fra le applicazioni ora allo studio si annoverano microlenti per *computer vision*, multiplatori di luce per comunicazioni ottiche e addirittura microimpianti ottici binari per correggere difetti visivi. Molte imprese sono impegnate nella ricerca sull'ottica binaria e alcune hanno incominciato a mettere a punto nuovi prodotti. L'ambito di questa tecnologia è così ampio che la maggior parte del suo potenziale deve essere ancora esplorata.

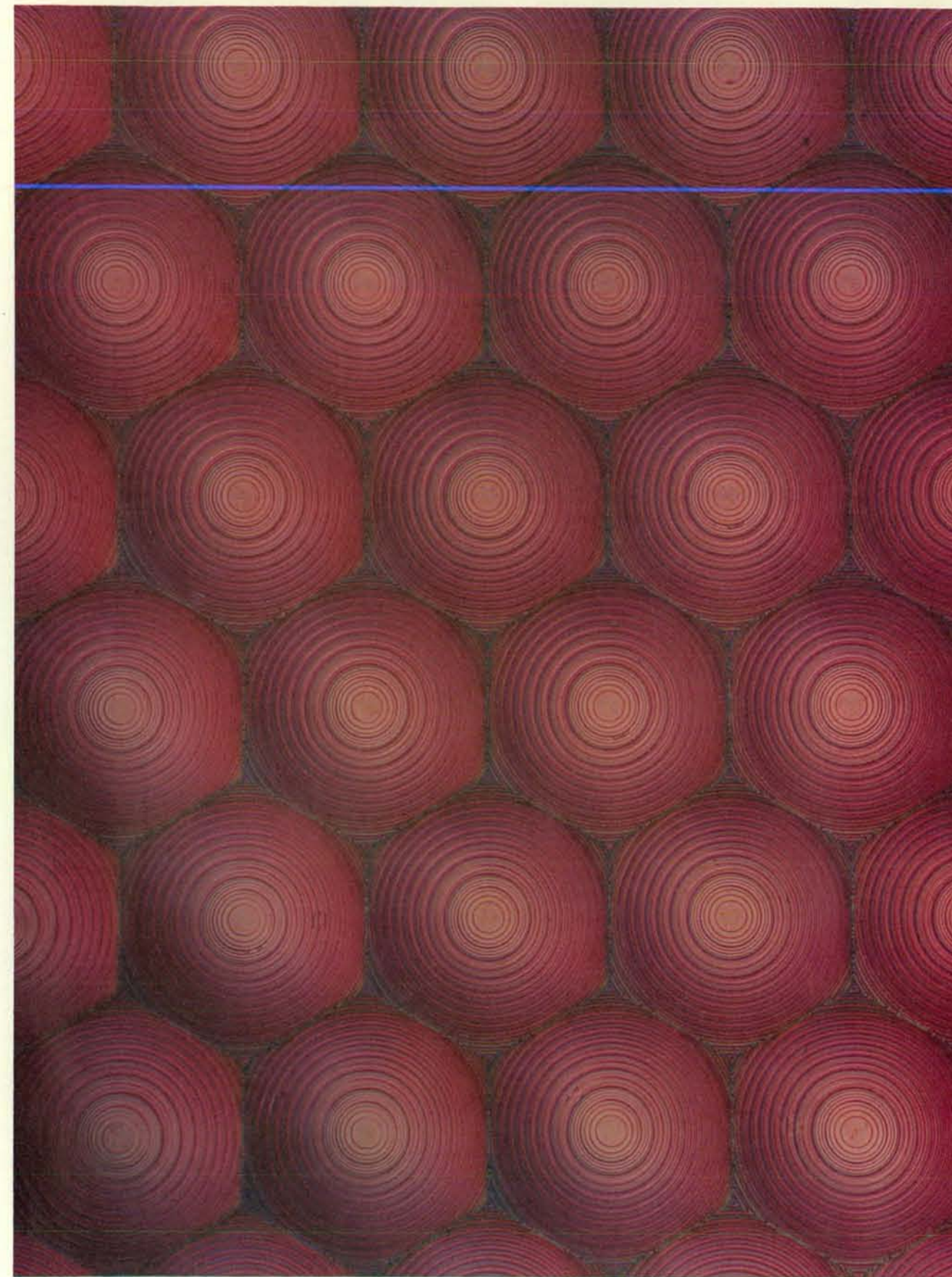
I dispositivi fondati sull'ottica binaria non sono soltanto realizzati con metodi del tutto diversi da quelli impiegati per produrre la maggior parte delle lenti, ma funzionano anche secondo un insieme di principi differenti: si basano infatti sulla diffrazione anziché sulla rifrazione. Una lente convenzionale fa convergere in un fuoco la luce proveniente da una sorgente lontana perché i raggi che attraversano la periferia della lente ne colpiscono la superficie a un angolo di incidenza maggiore e risultano perciò maggiormente deflessi rispetto a quelli che passano attraverso il centro. La deflessione è una conseguenza inevitabile del passaggio di luce tra sostanze aventi indici di rifrazione diversi. Viceversa, uno strumento basato sull'ottica binaria spezza il fronte d'onda della luce in arrivo in ogni punto della superficie della lente e lo ricostruisce sotto forma di un'onda che si muove nella direzione desiderata verso il punto focale.

La maggior parte dei sistemi ottici rifrattivi (una categoria che include quasi tutto fatta eccezione per i telescopi astronomici, che si basano sulla riflessione) soffre di un certo numero di imperfezioni intrinseche, come l'aberrazione sferica. La luce che attraversa parti diverse della lente converge in punti diversi lungo l'asse ottico, generando un'immagine sfocata. Molti progettisti devono spesso ricorrere a lenti di differenti materiali e curvature per poter compensare l'aberrazione.

Gli elementi ottici diffrattivi possono essere molto più sottili e leggeri di quelli rifrattivi poiché devono indurre variazioni dell'ordine di una frazione di lunghezza d'onda nella forma di un fronte d'onda ottico. Essi si basano sul principio di Huygens: ogni punto del fronte d'onda del fascio di luce manipolato agisce come una sorgente di onde sferiche, la cui interferenza costruttiva e distruttiva produce un nuovo fronte d'onda. La forma di questo fronte d'onda può essere alterata in modo che il percorso della luce venga deflesso, converga a un fuoco o manifesti qualunque altro comportamento scelto dal progettista.

Forse l'esempio più semplice di questo fenomeno è il reticolo a zone di Fresnel, che sfrutta la diffrazione per mettere a fuoco la luce monocromatica. I reticoli a zone consistono di una serie di anelli concentrici che bloccano la luce, separati da fenditure. La distanza tra queste ultime è scelta in modo che le onde luminose provenienti da tutte le fenditure interferiscano distruttivamente, tranne che nel punto focale del reticolo dove si rinforzano a vicenda e creano una macchia luminosa. Le tecniche di litografia e di incisione usate per costruire i circuiti integrati sono perfettamente adatte a realizzare una struttura come il reticolo a zone: le aree destinate alle fenditure vengono incise mentre il resto del materiale non viene toccato.

Anche se sono semplici da costruire, i reticoli a zone di Fresnel sono otticamente inefficienti poiché meno di metà della luce che li colpisce passa attraverso le fenditure. L'ottica binaria produce effetti simili con maggior efficienza, dal momento che si incidono parti diverse di un pezzo ottico grezzo a profondità differenti invece di limitarsi a bloccare la luce o a consentirne il passaggio.

La struttura ottica binaria più semplice è il prisma, che al microscopio appare come una serie di minutissimi scalini. Quando la luce colpisce la superficie del prisma, l'onda, sostanzialmente, viene spezzata in fronti d'onda secondari. Ogni fronte d'onda viene ritardato in proporzione allo spessore della struttura a scalini in quel punto (il massimo spessore è appena sufficiente per ritardare un

Questa schiera di lenti microscopiche è stata realizzata con le stesse tecniche usate per incidere circuiti integrati su substrati di silicio: ogni lente ha un diametro di 200 micrometri. Gli autori prevedono che tali schiere contribuiranno a rivoluzionare la visione artificiale focalizzando la luce su minutissimi fotorivelatori circondati da circuiti per l'elaborazione di immagini.

fronte di un'intera lunghezza d'onda). Quando questi fronti d'onda interferiscono, producono un nuovo fronte d'onda deflesso di un certo angolo rispetto a quello in arrivo. Così, di fatto, la luce viene deviata. Invece dei millimetri di vetro necessari per deflettere la luce in un prisma convenzionale, tuttavia, la versione ottica binaria richiede appena due micrometri di materiale.

Una lente è essenzialmente un prisma variabile, in grado di deflettere la luce a diversi angoli, a seconda di dove viene colpita dalla luce stessa. Così, la tecnica a scalini può essere estesa alla realizzazione di un dispositivo di messa a fuoco in ottica binaria. Un pezzo circolare viene inciso con una serie di strutture concentriche a scalini. A partire dal centro, ogni scalino è via via più alto e più vicino a quello adiacente. Scale più ripide deflettono la luce di angoli via via crescenti, in modo tale che essa, passando attraverso i bordi della lente, converga verso lo stesso punto al quale converge la luce che attraversa il centro. Scegliendo accuratamente la ripidità e la distanza degli scalini, si può eliminare completamente l'aberrazione sferica.

Sono possibili molti altri tipi di microstrutture a scalini. Lenti in cui gli scalini sono alti due o più lunghezze d'onda sono in grado di mettere a fuoco la luce in un intervallo più ampio di lunghezze d'onda e diffondono meno luce rispetto alle versioni con scalini alti una sola lunghezza d'onda. Altezze corrispondenti a frazioni di una lunghezza d'onda producono componenti binari che fungono da divisori o da combinatori di fascio.

Le attuali tecniche di produzione di circuiti integrati riescono a fabbricare configurazioni le cui dimensioni possono arrivare a mezzo micrometro; l'altezza delle strutture può essere controllata con la precisione di un nanometro. Così si possono realizzare dispositivi ottici binari mediante l'attuale tecnologia di fabbricazione dei semiconduttori, minimizzando i costi di avviamento.

Inoltre la realizzazione litografica di strutture a scalini è economica. Ogni stadio di fabbricazione raddoppia il numero degli scalini. Una prima fase di esposizione e di incisione ne produce due; ripetendo il procedimento con un'altra maschera avente caratteristiche di larghezza dimezzata e incidendo per metà della profondità precedente si ottengono quattro scalini; un terzo ciclo porta a otto scalini e così via. I dispositivi ottici binari non richiedono più di quattro cicli di incisione, e perciò sono più semplici da produrre rispetto ai circuiti integrati, che necessitano anche di una decina di cicli d'incisione alternati alle operazioni di impiantazione ionica e di deposizione di polisilicio, metalli e isolanti.

La semplicità delle fasi ripetute di mascheratura e di incisione rende possibile la fabbricazione di elementi complessi di ottica binaria a basso costo e in grande quantità, ma la tecnica impiegata è anche causa di un inconveniente significativo. Essa infatti non produce le superfici curve e angolate dell'ottica convenzionale, ma piuttosto una serie di scalini piatti, che approssima la forma voluta. Le deviazioni dalla forma ideale fanno diffondere la luce fuori dal cammino previsto. Perdite anche del 5 per cento alla superficie di ogni elemento ottico binario possono compromettere seriamente le prestazioni di treni ottici contenenti più di cinque o sei elementi. L'efficienza ottica - la quantità di luce indirizzata nella direzione voluta - può essere aumentata con approssimazioni più vicine alla forma ideale. Due scalini corrispondono, nel migliore dei casi, a un'efficienza del 41 per cento, quattro scalini all'81 per cento, otto scalini al 95 per cento e 16 al 99 per cento.

Benché i progressi compiuti negli anni settanta e ottanta nella tecnologia di costruzione dei circuiti integrati abbiano reso possibile l'ottica binaria, gli elementi ottici diffrattivi non sono nuovi: da decenni esistono progetti per molti diversi tipi di ottica diffrattiva. Fra di essi vi sono reticoli di diffrazione variabili, *kinoform* (lenti a gradini la cui superficie si riduce esattamente di una lunghezza d'onda della luce a ogni aumento di spessore della stessa quantità) e ologrammi. (Sebbene gli ologrammi più noti manipolino fronti d'onda ottici piani per produrre immagini, si sono anche realizzati ologrammi solo per deflettere con efficienza la luce a vari angoli, senza formare un'immagine.) Questi componenti diffrattivi non erano tuttavia largamente impiegati a causa della difficoltà di costruzione.

Sul finire degli anni settanta, un gruppo di ricercatori del Lincoln Laboratory al Massachusetts Institute of Technology incominciò a lavorare sull'ottica diffrattiva per manipolare la luce nei sensori radar a laser, che sfruttano la luce per rivelare oggetti più piccoli di quelli visibili con il radar convenzionale. Il radar a laser contiene una schiera lineare di rivelatori. Convogliare il fascio di ritorno ai rivelatori per mezzo di elementi ottici convenzionali, come per esempio specchi parzialmente riflettenti, si dimostrò poco pratico. Il treno ottico richiedeva aggiustamenti minuziosi e poco riproducibili e gran parte della luce veniva dispersa.

Fu allora progettato un mosaico diffrattivo in grado di dividere un singolo fascio laser in 12 fasci di ampiezza e fase determinate con precisione. Le tecniche di stampa di microcircuiti impiegate per costruire il mosaico potevano produrre soltanto due scalini, invece delle strutture a più scalini usate nell'ottica binaria più recente. Ciononostante, la luce che colpiva le creste sinuose della superficie del mosaico veniva diffratta in modo tale che la figura di interferenza risultante produceva 12 nuovi fasci, ciascuno dei quali si allontanava dalla superficie a un angolo diverso. Accoppiando con precisione i fasci alla schiera di rivelatori del ricevitore, l'elemento diffrattivo migliorava in misura significativa l'efficienza di rilevamento e la risoluzione del dispositivo.

Ulteriori ricerche vennero promosse dalla Defense Advanced Research Projects Agency nell'ambito degli sforzi compiuti dal Defense Department per sviluppare sensori per l'infrarosso a basso costo. Questa ricerca avrebbe infine condotto alla fabbricazione di microstrutture ottiche a più scalini.

Quasi contemporaneamente, i progettisti dell'attuale Hughes Danbury Optical Systems stavano lavorando all'applicazione di elementi olografici diffrattivi a specchi per telescopi di enormi dimensioni. Questi elementi erano necessari per deviare una piccola quantità di radiazione laser in un sensore che controllava la prestazione ottica del sistema.

Gli specchi erano troppo grandi per poter essere disposti su un banco ottico dove le loro superfici potessero essere rivestite di emulsione e gli ologrammi esposti con metodi convenzionali. Si decise, invece, di calcolare matematicamente le configurazioni di fase degli ologrammi desiderati e poi di ricostruirle e inciderle nelle superfici degli specchi usando tecniche microlitografiche mutuate dall'industria dei semiconduttori. Il successo di questo progetto ha condotto ad applicazioni di grande portata. Oltre a generare ologrammi, l'ottica binaria è anche in grado di produrre virtualmente qualsiasi trasformazione di un fronte d'onda ottico.

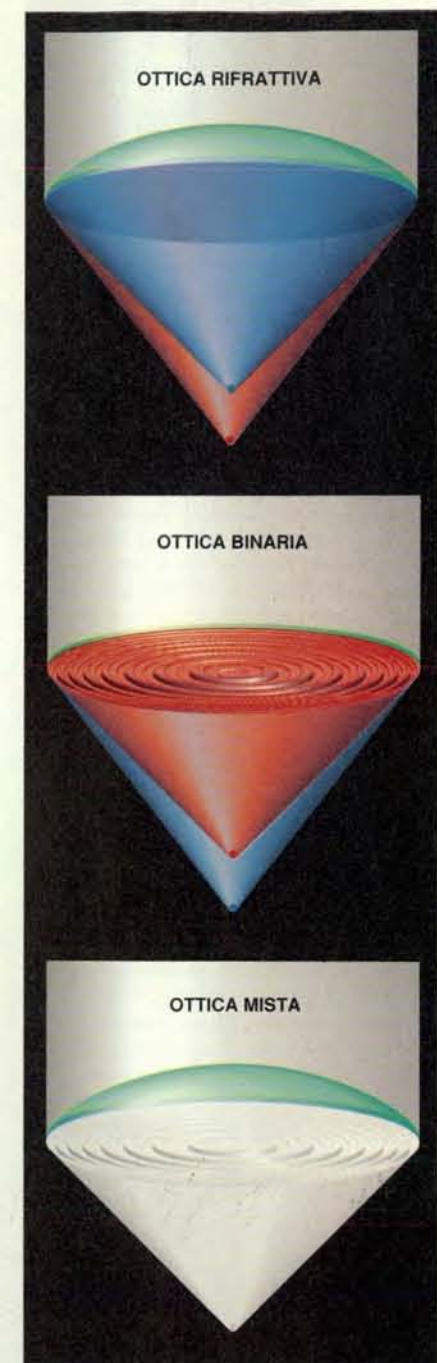
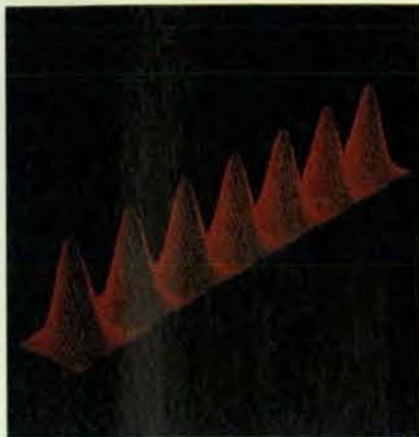
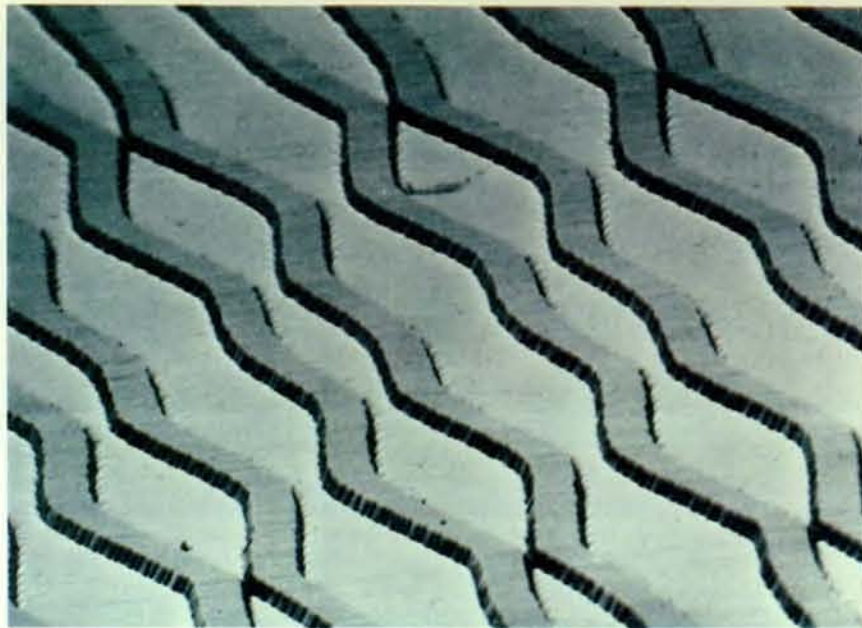
Innanzitutto i ricercatori della Hughes verificarono le proprie nuove capacità

La diffrazione è alla base dell'ottica binaria

Gli effetti di interferenza permettono all'ottica binaria di curvare e mettere a fuoco la luce raggiungendo gli stessi risultati dell'ottica rifrattiva convenzionale. Il più semplice esempio di diffrazione è l'esperimento con due fenditure (*a destra*), in cui le creste e i ventri delle onde luminose che attraversano due piccole aperture si combinano formando una figura di frange chiare e scure.

Secondo il principio di Huygens, ogni punto di un fronte d'onda ottico può essere considerato come una minutissima fenditura che emette una nuova onda sferica, la cui interferenza con le vicine crea il successivo insieme di creste e ventri. Quando tale fronte d'onda attraversa un dispositivo ottico binario (*in basso*), ogni sua parte subisce un ritardo proporzionale allo spessore della struttura a scalini del dispositivo. L'interferenza fra le diverse parti del fronte d'onda produce una nuova onda che si muove in una diversa direzione.

DIFFRAZIONE A DUE FENDITURE

DIFFRAZIONE A SCALINI

Le lenti convenzionali (*in alto*) focalizzano la luce blu più di quella rossa; questo fenomeno è detto aberrazione cromatica. Gli elementi ottici binari (*al centro*) mettono invece a fuoco la luce rossa più di quella blu. Combinando l'ottica convenzionale e quella binaria in un unico elemento (*in basso*), si può eliminare l'aberrazione cromatica in un intervallo limitato di lunghezze d'onda.

Il multiplatore ottico binario (*a destra*) scompone un fascio laser in sette fasci di uguale ampiezza (*sopra*), ciascuno deviato di un certo angolo rispetto alla luce incidente. Questo compito non può essere svolto dall'ottica convenzionale.

creando un componente ottico binario per possibili applicazioni in un sistema di riconoscimento di forme. Il dispositivo era ideato per rendere il sistema più sensibile ai dettagli della forma di un oggetto e meno sensibile alle sue dimensioni o all'orientazione rotazionale.

Le proprietà ottiche del dispositivo trasformano cerchi concentrici in linee rette parallele. Pertanto qualunque oggetto centrato nel campo visivo viene sostanzialmente affettato in sottili anelli, ciascuno dei quali è trasformato in una banda rettangolare. Quando l'immagine di un oggetto viene ruotata attorno al centro del campo, ogni caratteristica distintiva scorre da una parte all'altra della banda rettangolare. La maggior parte dei sistemi per il riconoscimento di forme può identificare tali caratteristiche traslate molto più facilmente di quanto possa fare con caratteristiche ruotate.

L'idea alla base di un simile dispositivo non è nuova, ma nessun ottico potrebbe realizzare il componente necessario ricorrendo alle tecniche convenzionali di molatura e levigatura. I tentativi per produrlo mediante ologrammi sintetici (ossia generati con il calcolatore) si rivelarono scomodi e otticamente inefficienti: la versione ottica binaria superò tutte le altre. Da allora sono state progettate lenti per correggere l'astigmatismo, l'aberrazione sferica e un buon numero di altri difetti ottici.

Uno dei pochi ostacoli inevitabili all'uso di elementi ottici diffrattivi è che questi sono altamente dispersivi, ossia tendono a dividere la luce nei colori componenti. Tuttavia in alcuni casi si è scoperto come trasformare questa debolezza in un punto di forza. Si possono combinare elementi ottici binari con lenti convenzionali per ridurre l'aberrazione cromatica, che deriva dal carattere dispersivo del vetro e degli altri materiali rifrattivi.

Nella progettazione tradizionale delle lenti, si minimizzano le aberrazioni cromatiche saldando insieme lenti di materiali con diverse proprietà dispersive. Questa tecnica fa aumentare il costo e la complessità del prodotto finito. La dispersione dell'ottica binaria, per fortuna, è di segno opposto rispetto a quella della maggior parte dei vetri; così, su un limitato intervallo di lunghezze d'onda, l'aberrazione cromatica di una lente binaria elimina quella di una lente convenzionale nella quale sia stata incisa. I principi della correzione diffrattiva erano noti da tempo, ma non erano applicabili nella pratica. Gli ologrammi diffrattivi sintetici erano ritenuti troppo inefficienti, gli ologrammi ottici erano troppo difficili da produrre, e la pellicola olografica o i materiali gelatinosi erano troppo instabili per un lavoro di alta precisione.

L'ottica binaria supera queste limitazioni consentendo l'incisione di elementi diffrattivi ad alta efficienza direttamente in un materiale ottico selezionato. Di fatto si possono sintetizzare nuovi tipi di vetro la cui capacità di deflettere la luce vari opportunamente al variare della lunghezza d'onda. Questa ulteriore flessibilità può rendere superfluo lavorare con materiali ottici costosi o difficili da trattare, riduce il numero di elementi ottici in un sistema e ne migliora in generale le prestazioni.

La trasformazione ottica converte cerchi concentrici (*a sinistra*) in linee rette (*a destra*). Il passaggio dal sistema di coordinate polari a quello cartesiano può perfezionare i dispositivi per il riconoscimento di forme, rendendoli meno sensibili alla rotazione del bersaglio. La figura al centro è una mappa a curve di livello dell'elemento ottico che forma l'immagine.

Presso il Lincoln Laboratory fu progettata una lente ottica binaria che venne incisa sulla faccia piana di una lente piano-convessa (una lente dotata di una faccia piana e di una curva). La lente ibrida, in silice fusa in grado di trasmettere la radiazione ultravioletta, non presentava aberrazione né cromatica né sferica in un intervallo di lunghezze d'onda compreso fra 246 e 250 nanometri. L'aberrazione cromatica in quella parte dello spettro è particolarmente intensa: una lente non corretta ha una larghezza di banda oltre mille volte inferiore a quella dell'ibrido ottico convenzionale-binario, nei limiti imposti dalla diffrazione.

Poiché gli strumenti ottici binari e i chip semiconduttori vengono costruiti con le stesse apparecchiature, sarà possibile incidere i componenti ottici direttamente sui circuiti integrati. Si potranno disporre su un chip schiere di microlenti o addirittura unità di elaborazione optoelettroniche integrate. È proprio da applicazioni di questo genere che nasceranno nuove prospettive per sistemi di sensori «intelligenti» e ultracompatti.

Come primo passo in questa direzione, si sono già realizzate lenti ottiche binarie di diametro prossimo a quello di un capello umano (da 50 a 100 micrometri). Queste micro-ottiche binarie possono essere costruite singolarmente o in schiere di decine di migliaia di lenti per centimetro quadrato.

Le lenti non si limiteranno a mettere a fuoco la luce su schiere di minutissimi fotorivelatori elettronici; sosterranno invece un ruolo importantissimo nel dar forma all'architettura dei futuri sistemi di elaborazione di immagini. Le attuali architetture per la visione artificiale si basano su schiere di rivelatori che generano immagini simili a quelle fotografiche, dalle quali con impegnativi calcoli si estraggono caratteristiche come bordi e superfici e, infine, la descrizione degli oggetti in una scena. Lenti microscopiche capaci di mettere a fuoco la luce incidente su un rivelatore lasceranno spazio per elementi di elaborazione connessi direttamente a ciascun rivelatore. Si potranno costruire schiere di rivelatori che imitino il processo visivo degli organismi viventi, che, finora, si è dimostrato superiore a tutte le versioni artificiali (si veda l'articolo *La retina di silicio* di Misha A. Mahowald e Carver Mead in «Le Scienze» n. 275, luglio 1991).

Questo miglioramento offrirà la possibilità di costruire sensori per immagini capaci di adattarsi alla variazione dei livelli di luce, di rilevare il movimento e di determinare automaticamente la posizione del contorno di un oggetto in un ambiente. Uno di noi (Veldkamp) ha coniato il termine *amacronics* (amacronici) per questi circuiti, perché simulano le funzioni di rilevamento del moto, di accentuazione dei contorni e di riduzione del campo dinamico che sono proprie delle cellule amacrine della retina.

Il modello biologico può essere emulato con sistemi ancora più complessi. Elementi addizionali di ottica binaria possono interconnettere le cellule di schiere di fotorivelatori disposte in strati come le cellule della retina. Una serie di diodi a emissione luminosa e di dispositivi micro-ottici integrati potrebbe trasmettere l'immagine di basso livello a un altro chip, i cui sensori potrebbero essere sensibili, per esempio, all'orientazione o alle caratteristiche strutturali dell'immagine. Un insieme di circuiti per l'elaborazione di immagini potrebbe alla fine trasmettere una descrizione prevalentemente simbolica di una scena in modo analogo a quello con cui la corteccia visiva comunica la propria informazione ad altre aree del cervello.

Oltre a migliorare la visione artificiale, gli strumenti amacronici saranno utili per ridurre i requisiti, in termini di larghezza di banda, per la trasmissione di immagini. Invece di inviare un'immagine completa, questi sistemi potranno estrarre e trasmettere informazioni chiave in base alle quali il ricevitore possa ricostruire l'immagine. Numerose società giapponesi hanno già avviato lo sviluppo di sistemi amacronici per applicazioni nel campo dei videotelefoni.

La rivoluzione microelettronica deve gran parte del proprio successo all'ottica di alta qualità. Senza i sofisticati sistemi ottici che proiettano immagini ridotte di maschere circuitali su substrati di silicio, la produzione di massa a basso costo dei microprocessori sarebbe stata impossibile. Ora, con lo sviluppo dell'ottica binaria, la microelettronica sembra ripagare il proprio debito in un modo che pochi avrebbero potuto prevedere. I calcolatori e le moderne apparecchiature per la lavorazione dei semiconduttori vengono impiegati per costruire elementi ottici che un tempo si ritenevano impossibili.

In effetti, così come oggi i progettisti di circuiti siedono a un terminale e maneggiano strumenti di costruzione elettronica a blocchi, anche i progettisti di ottica binaria potranno un giorno sintetizzare sistemi optoelettronici agendo semplicemente su una tastiera. Creando progetti che impieghino sia l'ottica binaria sia quella convenzionale, si può spesso dimezzare il numero di elementi di un sistema complesso.

Trasformare questa prospettiva in realtà richiede lo sviluppo di tecniche migliori per i progetti di ottica binaria, forse attraverso la rivisitazione di metodi che furono abbandonati anni fa in quanto a quel tempo l'ottica diffrattiva generalizzata, di alta qualità, era ritenuta priva di applicazioni pratiche. Abbiamo trovato che vecchi libri e documenti tecnici dimenticati possono costituire una ricca sorgente di idee per progetti che oggi sono realizzabili. Naturalmente l'ottica convenzionale e i metodi di costruzione verificati nel tempo non stanno per scomparire, né debbono farlo. L'ottica basata sulla riflessione e sulla rifrazione continuerà a essere utilizzata nelle applicazioni in cui funziona meglio.

L'ottica binaria offre una gamma così ampia di possibilità di manipolazione della luce che ai progettisti si impone la sfida di imparare in quali occasioni non convenga usare la nuova tecnica e in quali, invece, sia bene sfruttarla. Dato che le nuove lenti si stanno affermando in applicazioni che vanno dall'ottica convenzionale per la produzione di immagini alle interconnessioni con circuiti integrati e alla visione artificiale, i progettisti si vengono a trovare in una situazione insolita: limitati non dai materiali o dalle tecniche di costruzione, ma soltanto dalle proprie capacità creative.

Una lente binaria sulla superficie della cornea potrebbe correggere difetti visivi come la miopia, l'ipermetropia o l'astigmatismo. La lente qui illustrata è stata incisa con impulsi laser ultravioletti in una cornea prelevata da un cadavere.

BIBLIOGRAFIA

LEE WAI-HON, *High Efficiency Multiple Beam Gratings* in «Applied Optics», 18, n. 13, 1 luglio 1979.

KORONKEVICH VOLDEMAR P., *Computer Synthesis of Diffraction Optical Elements* in *Optical Processing and Computing*, Academic Press, 1989.

VELDKAMP WILFRID, *Binary Optics* in *McGraw-Hill Yearbook of Science & Technology: 1990*, McGraw-Hill, 1989.

MAHOWALD MISHA A. e MEAD CARVER, *La retina di silicio* in «Le Scienze» n. 275, luglio 1991.

Heisenberg, l'indeterminazione e la rivoluzione quantistica

Spinto dall'ambizione e da un fiero spirito di competizione, Werner Heisenberg formulò uno dei principi più noti della scienza e fu, appena trentaduenne, tra i più giovani scienziati a ricevere il premio Nobel

di David C. Cassidy

Fra le tante conquiste della scienza di questo secolo, la più fondamentale è forse la meccanica quantistica. Formulata da un manipolo di fisici europei di grande talento, la scienza dell'atomo ci ha obbligato ad alcune profonde e controverse modificazioni delle nostre visioni sulla natura. A seconda di come si compiono le osservazioni, la materia si può presentare sotto forma di onde o di particelle e causa ed effetto non sono più intimamente legati. Questa interpretazione della meccanica quantistica (le regole su come e quando applicarla e ciò che essa ci dice del mondo fisico) venne formulata a Copenaghen nel 1927. A causa della diffusione avviata dai suoi fondatori e del successo sbalorditivo nelle applicazioni, già negli anni trenta l'interpretazione di Copenaghen aveva raggiunto la preminenza di cui gode oggi. Ma un'interpretazione non è che un'interpretazione. Le sue origini, il sostegno fornitole e la sua accettazione possono essere, sotto diversi importanti aspetti, tanto un risultato di circostanze storiche e di preferenze individuali quanto un effetto della sua validità scientifica.

Il ruolo che l'indole dell'uomo ha nella scienza emerge con evidenza dalle vicende di uno dei principali ideatori e più attivi sostenitori dell'interpretazione di Copenaghen, Werner Karl Heisenberg. Fu nel 1927 che questi, venticinquenne assistente di Niels Bohr, diede il suo più celebre contributo alla fisica: il principio di indeterminazione, che costituisce un tassello fondamentale nel mosaico dell'interpretazione di Copenaghen. Questo principio, come del resto tutta l'interpretazione, scaturì dalla ricerca di un legame coerente tra il mondo quotidiano del laboratorio e lo strano mondo microscopico dell'atomo.

In breve, il principio d'indeterminazione asserisce che la misurazione simultanea di due variabili coniugate, come la posizione e la quantità di moto (o momento lineare) di una particella, si può effettuare solo con precisione limitata. Quanto più precisa è la misurazione della posizione, tanto più imprecisa è quella della quantità di moto e viceversa. Nel caso limite, una precisione assoluta per una delle variabili comporterebbe un'imprecisione assoluta per l'altra.

Questa indeterminazione non è attribuibile all'errore sperimentale, ma è una conseguenza fondamentale delle equazioni quantistiche ed è caratteristica di qualsiasi esperimento quantistico. Inoltre, dichiarò Heisenberg, finché la meccanica quantistica resterà valida, il principio d'indeterminazione non sarà mai violato. Era la prima volta dalla rivoluzione galileiana che un fisico eminente proclamava l'esistenza di limiti alla comprensione scientifica.

Insieme con le idee di luminari come Bohr e Max Born, il principio d'indeterminazione di Heisenberg costituì il sistema perfettamente autoconsistente dell'interpretazione di Copenaghen. Davanti ai più eminenti fisici quantistici del mondo riuniti a Bruxelles nell'ottobre del 1927 per la quinta delle famose Conferenze Solvay sulla fisica teorica, Heisenberg e Born dichiararono che questo sistema era completo e irrevocabile. Il congresso seguiva di poche settimane la nomina di Heisenberg alla cattedra di fisica teorica dell'Università di Lipsia. Il venticinquenne Heisenberg era così il più giovane dei professori ordinari della Germania.

La giovanissima età che Heisenberg aveva quando fornì il suo contributo più importante è indicativa di una caratteristica tipica di questa e delle altre sue prime ricerche: il desiderio quasi insaziabile di ottenere onori accademici e di distinguersi sempre come il migliore in tutto ciò che faceva. Non sorprende che questa sua brama possa farsi risalire in buona parte alle tradizioni familiari.

Gli Heisenberg erano una famiglia ambiziosa e di grande cultura dell'alta borghesia tedesca. L'unificazione della Germania dovuta a Otto von Bismarck nella seconda metà dell'Ottocento e il conseguente forte sviluppo dell'economia avevano prodotto una cospicua richiesta di burocrati, diplomatici, giudici, avvocati e dirigenti. Di conseguenza, le università e le scuole della giovane nazione acquisirono d'improvviso grande importanza. Crebbero di pari passo il prestigio e la remunerazione degli accademici e dei loro discepoli migliori.

Sia il padre di Werner, August, sia il nonno materno, Nikolaus Wecklein, erano di umili origini ed erano saliti al vertice dell'alta borghesia tedesca grazie al successo accademico. Wecklein era direttore di un famoso ginnasio di Monaco e nel 1910 August divenne professore di filologia bizantina all'Università della capitale bavarese. Entrambi presero poi moglie all'interno del loro nuovo ceto.

Fin dalla sua nascita, nel 1901, i familiari avevano deciso che anche Werner avrebbe raggiunto un'alta posizione sociale grazie al sapere. Ritenendo che la competizione ne avrebbe favorito il successo accademico, August alimentò un'intensa rivalità tra Werner e il fratello maggiore Erwin. Nel corso degli anni i due fratelli ingaggiarono lotte furibonde e la loro rivalità culminò un giorno in una rissa sanguinosa, in cui si batterono a colpi di sedia. Divenuti adulti, andarono ognuno per la sua strada. Erwin si trasferì a Berlino e diventò chimico e, tranne che per occasionali riunioni di famiglia, i fratelli ebbero pochi contatti.

L'ambizione di Werner di raggiungere il vertice è particolarmente evidente nel periodo che va dal luglio 1925, quando, con i colleghi Born e Pascual Jor-

Werner Heisenberg diede i contributi più importanti alla fisica fra i venti e i trent'anni. Questa fotografia fu scattata intorno al 1924 all'Università di Gottinga, dove Heisenberg ottenne l'abilitazione a ricoprire una cattedra universitaria.

dan, formulò una descrizione matematica della meccanica quantistica, al febbraio 1927, quando ricavò le relazioni di indeterminazione. Ciò che rese così importanti le conseguenze di tanta ambizione fu la concomitanza di due eventi.

In primo luogo, nei paesi di lingua tedesca dell'Europa centrale si liberarono d'un tratto parecchie cattedre di fisica teorica. Questi posti rappresentavano un'ottima opportunità per un accademico ambizioso come Heisenberg, che aveva già ottenuto la libera docenza all'Università di Gottinga ed era quindi abilitato a ricoprire una cattedra.

L'altra circostanza, forse ancora più importante, fu la comparsa di una seconda descrizione matematica della meccanica quantistica. Nel 1925 Heisenberg e colleghi avevano sviluppato per la meccanica quantistica un formalismo basato sulla matematica astratta del calcolo delle matrici. Agli occhi degli autori, questa «meccanica delle matrici» secondava la loro inclinazione a considerare le sole grandezze osservabili in laboratorio. Essi sostenevano l'esistenza di entità essenziali, come salti quantici e discontinuità all'interno dell'atomo, e respingevano l'idea di modelli atomici *anschaulich*, cioè visualizzabili.

Erwin Schrödinger, trentanovenne fisico viennese che allora lavorava a Zurigo, affrontò il rompicapo della fisica atomica da una prospettiva e con fini affatto diversi. In una serie di memorie pubblicate nella prima metà del 1926, Schrödinger presentò un'equazione d'onda quantistica basata su una proposta avanzata dal francese Louis-Victor de Broglie nella sua tesi di dottorato. L'idea, accolta favorevolmente anche da Einstein, era che tutta la materia in moto si poteva considerare composta da onde. Schrödinger sfruttò questa ipotesi per sostenere che le «onde di materia» dell'elettrone instauravano all'interno dell'atomo modi vibratori armonici. Questi modi sostituiscono gli stati atomici stazionari della teoria delle matrici e i salti quantici discontinui diventano transizioni continue da un modo armonico all'altro. Se fosse stata vera, la formulazione di Schrödinger avrebbe vanificato le entità essenziali della meccanica delle matrici di Heisenberg.

La maggioranza dei fisici accolse con favore l'impostazione di Schrödinger, più consueta, e non si sbilanciò sulla sua interpretazione. La situazione subì una brusca svolta nel maggio 1926, quando Schrödinger pubblicò una dimostrazione dell'equivalenza dei due formalismi matematici. Heisenberg e i suoi sostenitori serrarono subito i ranghi e su entrambi i fronti la polemica assunse toni sempre più accesi.

Schrödinger non contribuì a placare gli animi. Nella sua memoria sull'equivalenza egli attribuiva al proprio schema un valore superiore all'altro. In una famosa nota a piè di pagina affermò: «Che io sappia [tra il lavoro di Heisenberg e il mio] non esiste alcun legame di parentela. Naturalmente ero al corrente della sua teoria, ma mi sentivo scoraggiato, se non addirittura respinto, dai metodi dell'algebra trascendente e dalla mancanza di *Anschaulichkeit* [visualizzabilità].»

Heisenberg rispose a tono in una lettera al suo collega e amico Wolfgang Pauli: «Più rifletto sugli aspetti fisici della teoria di Schrödinger, più li trovo irritanti... Ciò che Schrödinger scrive sulla visualizzabilità della sua teoria "probabilmente non è affatto corretto" [un'eco di Bohr], in altre parole è *Mist* [letame].» L'unico vantaggio del metodo di Schrödinger, sostenne in pubblico Heisenberg, era che consentiva un semplice calcolo della probabilità delle transizioni atomiche, cioè della probabilità dei salti quantici, da introdurre nelle matrici della meccanica quantistica. Pauli si disse d'accordo.

Esaminando meglio la cronologia di queste osservazioni, si vede che non fu l'equivalenza a provocare la contesa (Pauli l'aveva dimostrata in privato un mese prima senza tanto chiasso), bensì ciò che ne ricavarono le due parti. Heisenberg e gli altri della scuola delle matrici avevano dedicato tutta la loro carriera a misurarsi con proprietà della natura che credevano esistessero, incorporate nella loro descrizione matriciale. Avevano puntato il loro futuro su questa impostazione. Schrödinger aveva impegnato la propria reputazione nel tentativo di eliminare l'apparente irrazionalità delle discontinuità e dei salti quantici, sostituendola con una resuscitata fisica dei moti ondulatori continui, causali e razionali. Nessuna delle due parti era disposta a concedere all'altra la vittoria e - conseguenza probabile - il predominio professionale. D'un tratto furono messi in discussione la natura e il futuro sviluppo della meccanica quantistica.

La disputa contribuì a inasprire le ambizioni di carriera di Heisenberg. Solo alcune settimane prima che Schrödinger pubblicasse la dimostrazione di equivalenza, Heisenberg aveva rifiutato un posto di professore a Lipsia, preferendo un posto come assistente di Bohr a Copenaghen. Wecklein, il nonno di Werner, corse incredulo a Copenaghen per tentare di dissuadere il nipote dall'assumere quell'incarico proprio nel momento in cui compariva l'articolo di Schrödinger. Le rinnovate pressioni di Wecklein e la sfida lanciata da Schrödinger ai fondamenti della fisica delle matrici spinsero Heisenberg a raddoppiare gli sforzi per produrre un lavoro di qualità tale da farlo accogliere nei circoli specialistici e da procurargli alla fine qualche altra cattedra libera.

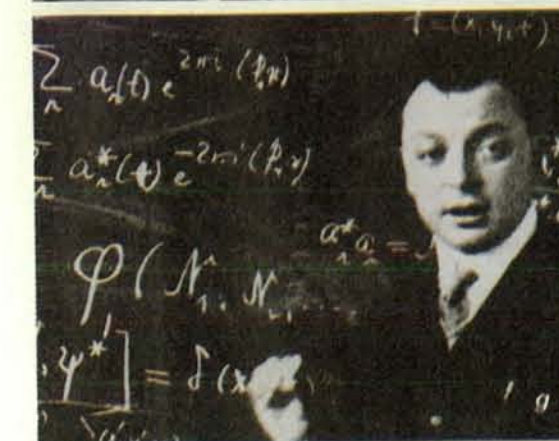
Ma nel 1926 almeno tre eventi gli fecero capire quanto fosse ampio il divario intellettuale tra le sue idee e il punto di vista di Schrödinger. Prima ci furono le conferenze sulla nuova fisica tenute da quest'ultimo a Monaco alla fine di luglio. Davanti al folto pubblico, il giovane Heisenberg sostenne che la teoria di Schrödinger non spiegava diversi fenomeni; tuttavia non riuscì a convincere nessuno e lasciò la conferenza scoraggiato. Poi, al convegno autunnale degli scienziati e medici tedeschi, Heisenberg constatò lo schiacciante favore, a suo modo di vedere mal riposto, di cui godevano le idee di Schrödinger.

Infine, nell'ottobre 1926, a Copenaghen, si svolse il dibattito, intenso ma in ultima analisi non decisivo, tra Bohr e Schrödinger. L'esito fu che nessuna delle interpretazioni dei due formalismi della meccanica quantistica venne riconosciuta del tutto accettabile. Chiunque, da una parte o dall'altra, avesse trovato un'interpretazione soddisfacente avrebbe potuto, sembra abbia detto Bohr, realizzare i propri «auspici» per una fisica del futuro.

Sotto la spinta di queste varie motivazioni personali, professionali e scientifiche, nel febbraio del 1927 Heisenberg credette di essersi improvvisamente imbattuto nell'interpretazione cercata: il principio d'indeterminazione. Il percorso intellettuale che lo portò a questa idea tra la fine del 1926 e l'inizio del 1927 partiva dalle ricerche dei suoi colleghi più stretti, specie Jordan e Paul Adrien Maurice Dirac, che insieme formularono la «teoria delle trasformazioni», la quale univa meccanica ondulatoria e meccanica matriciale. L'obiettivo di Heisenberg e alleati era quindi di scoprire un modo irrefutabile per incorporare la discontinuità nel formalismo di Dirac e Jordan.

Una spinta importante a favore della nuova interpretazione fu data da Pauli. Nella lettera del 19 ottobre 1926, in cui informava Heisenberg della dispo-

A esercitare le prime influenze sulla vita di Heisenberg furono il nonno Nikolaus Wecklein (*a*) e il padre August, qui con la moglie Anna e i figli Erwin (*in piedi*) e Werner (*b*). Furono i genitori ad accendere nei due ragazzi l'ambizione accademica. Heisenberg studiò sotto la guida di Niels Bohr (*c*), col quale formulò poi l'interpretazione di Copenaghen. Uno dei primi rivali di Heisenberg fu Erwin Schrödinger (*d*), il cui formalismo ondulatorio si oppose alla meccanica delle matrici proposta da Heisenberg con Max Born (*e*) e Pascual Jordan (*f*, *a destra*). Wolfgang Pauli (*g*) fu di grande aiuto a Heisenberg nel formulare nel 1927 il principio d'indeterminazione. Nel 1929 Heisenberg tenne una serie di conferenze per diffondere lo «spirito di Copenaghen», recandosi negli Stati Uniti, in Giappone, in Cina e infine in India (*h*).

La giovinezza di Werner Heisenberg

In una lunga lettera scritta a Wolfgang Pauli nel 1927, Heisenberg ricavò le relazioni d'indeterminazione per p e q, dove $p_1 = \sqrt{2}\Delta p$ e $q_1 = \sqrt{2}\Delta q$. Questo passo costituì il punto di partenza per la celebre memoria sul principio d'indeterminazione.

nibilità di una cattedra a Lipsia, Pauli applicò gli stati atomici stazionari alla precedente analisi di Born delle onde elettroniche libere. Egli trovò che per la quantità di moto p e la posizione q di un elettrone in un atomo si dovevano scegliere variabili continue, ma il loro comportamento quantistico manifestava un «punto oscuro»: «Si deve supporre che le p siano *controllate* e le q *non controllate*. Si possono cioè calcolare solo le probabilità di determinate variazioni delle p per condizioni iniziali assegnate e *mediate su tutti i valori possibili delle* q.» Non si può quindi parlare di una «determinata "traiettoria" della particella», scriveva Pauli, né «si può cercare simultaneamente il valore di p e il valore di q».

Heisenberg rispose di essersi «*molto* entusiasmato» per la lettera di Pauli e per il punto oscuro, su cui rifletté di frequente durante i mesi successivi. L'entusiasmo di Heisenberg culminò in una lettera di 14 pagine scritta a Pauli il 23 febbraio 1927 in cui esponeva quasi tutti i punti essenziali di una memoria che presentò dopo un mese esatto, intitolata *Sul contenuto* anschaulich *della cinematica e della meccanica teoriche quantistiche*: era la memoria in cui si introduceva il principio d'indeterminazione.

Avendo ricavato le relazioni di indeterminazione sia per via matematica sia attraverso esperimenti concettuali, Heisenberg riteneva che l'accordo tra i due metodi dimostrasse la validità universale dell'indeterminazione. La dimostrazione matematica prendeva le mosse da una funzione d'onda concepita come una curva a campana, cioè una densità di probabilità gaussiana, per la variabile q. L'errore relativo alla conoscenza del valore esatto di q (cioè la deviazione standard) è Δq. Sfruttando il formalismo sviluppato da Dirac e Jordan, Heisenberg trasformò la densità di probabilità gaussiana nella variabile coniugata p.

Nel far questo, scoprì che, per ragioni matematiche, le deviazioni standard delle due distribuzioni - cioè le imprecisioni sui valori di q e di p - sono tra loro inversamente proporzionali. Questa reciprocità può essere precisata ed espressa con la relazione:

$$\Delta p \cdot \Delta q \geq \frac{h}{4\pi},$$

dove h è la costante di Planck. Heisenberg dimostrò poi che questo non è solo un risultato astratto, ma è compatibile con qualunque esperimento immaginabile riguardante la misurazione simultanea di coppie di variabili coniugate, come posizione e quantità di moto o energia e tempo.

La compatibilità con gli esperimenti si basava tuttavia su svariate novità che Heisenberg aveva introdotto allo scopo di incorporare la discontinuità e le particelle. Tra queste vi era la ridefinizione del termine *anschaulich* nel titolo: esso non significava più visualizzabile o intuitivo, bensì «fisico» o significativo sotto il profilo empirico. Questa modificazione era intesa a parare la critica di Schrödinger che una fisica delle particelle discontinua è essenzialmente irrazionale e *unanschaulich*. Questa ridefinizione era strettamente legata alla ridefinizione di certi concetti classici come posizione, velocità e traiettoria di una particella atomica in termini delle operazioni sperimentali effettuate per misurare le grandezze suddette: una forma di operazionismo. Solo ciò che il fisico può misurare ha un significato reale; e in queste misurazioni si manifestano sempre le relazioni d'indeterminazione.

Per il giovane Heisenberg il principio d'indeterminazione costituiva il culmine e il completamento della rivoluzione quantistica, una rivoluzione che incorporava il suo apporto a fondamenti che lui stesso aveva concorso a costruire. E come per tacitare qualunque obiezione a questo punto di vista, concludeva la versione pubblicata della sua memoria con parecchie affermazioni che andavano molto al di là della matematica e dell'esperimento concettuale. Con la teoria delle trasformazioni di Dirac e Jordan, dichiarò Heisenberg, il formalismo quantistico è completo e inalterabile: le relazioni d'indeterminazione sono vere e irrefutabili poiché sono una conseguenza diretta del formalismo. Tutte le osservazioni passate e future dei fenomeni atomici ricadono quindi sotto questa interpretazione.

Inoltre, sosteneva, benché la fisica quantistica contenga un elemento statistico fondamentale, questo elemento non è una proprietà intrinseca della natura: esso si manifesta a causa della perturbazione provocata dallo scienziato nel tentativo di osservare la natura. Infine Heisenberg faceva la sua prima affermazione esplicita su una delle conseguenze più profonde dell'indeterminazione: una sfida alla causalità.

Il principio di causalità postula che ogni effetto sia preceduto da un'unica causa. Per oltre un secolo questa idea era stata a fondamento pressoché di tutte le forme di ricerca razionale. Al matematico francese Laplace si attribuisce la definizione forse più semplice di causalità applicata alla meccanica newtoniana: se la posizione e la quantità di moto di una particella sono noti con precisione in un istante dato, allora, conoscendo tutte le forze agenti sulla particella, il suo moto è completamente determinato in tutti gli istanti successivi dalle equazioni della meccanica.

Il principio d'indeterminazione, affermava Heisenberg, nega tutto ciò: «Nella formulazione rigorosa della legge di causalità, "se conosciamo il presente possiamo calcolare il futuro", non è sbagliata la conclusione, bensì la premessa.» I valori iniziali della quantità di moto e della posizione non possono essere misurati simultaneamente con precisione assoluta. Quindi per ogni istante futuro si può calcolare soltanto un inter-

vallo di valori possibili della posizione e della quantità di moto della particella. Il moto effettivo della particella indicherà una sola possibilità. Il legame causale tra presente e futuro è perduto e le leggi e le previsioni della meccanica quantistica assumono una natura puramente probabilistica, ossia statistica.

La memoria di Heisenberg sul principio d'indeterminazione era di vasta portata e di grande profondità sotto quasi tutti gli aspetti. Oltre a soddisfare così bene gli scopi dell'autore, l'articolo era consono al suo carattere. Quando il suo mentore Bohr gli fece osservare un errore nel ragionamento, Heisenberg difese con testardaggine la propria posizione, fino a che nella primavera del 1927 la disputa degenerò in quelle che Heisenberg chiamò «gravi incomprensioni personali». L'errore riguardava l'eccessiva fiducia che Heisenberg accordava alla discontinuità e alle caratteristiche corpuscolari dei quanti di luce in uno dei suoi fondamentali esperimenti concettuali, il cosiddetto microscopio a raggi gamma.

Bohr, tornando all'istituto da una vacanza sulla neve, trovò la memoria di Heisenberg già in bozza. Inviando, su richiesta dell'autore, il lavoro ad Einstein, Bohr si lamentò in privato con lui: l'impostazione di Heisenberg nel suo complesso era troppo limitata e l'esperimento del microscopio a raggi gamma era del tutto sbagliato, benché il risultato fosse corretto. Per Bohr le relazioni di indeterminazione non scaturivano soltanto dal formalismo, dalla ridefinizione di concetti fondamentali e dalla preminenza della discontinuità e delle particelle sulle onde continue. Erano di cruciale importanza anche il dualismo onda-particella e, nel microscopio a raggi gamma, la diffusione delle onde luminose provocata dagli elettroni nella lente del microscopio.

Le rappresentazioni ondulatoria e corpuscolare erano «complementari», mutuamente esclusive, ma congiuntamente essenziali. Bohr sosteneva che per analizzare l'esperimento lo sperimentatore doveva scegliere o la rappresentazione ondulatoria o quella corpuscolare. Il prezzo imposto da questa preferenza si esplicava in una limitazione di ciò che è possibile imparare dall'esperimento e questa limitazione era rappresentata dalle relazioni d'indeterminazione. Il ragionamento di Heisenberg era soltanto un caso particolare di quella che ora Bohr chiamava complementarità.

Heisenberg dissentì con veemenza e insistette sul valore fondamentale delle particelle e della discontinuità, rifiutando, nonostante il consiglio di Bohr, di ritirare l'articolo, che nel frattempo era stato mandato alle stampe. Heisenberg non tollerava il vasto impiego di onde o di concetti di meccanica ondulatoria né poteva rinunciare alla pubblicazione di quello che era il suo contributo più importante al dibattito sull'interpretazione. La conseguente disputa con Bohr divenne così aspra che, a quanto si dice, durante una riunione Werner scoppiò in lacrime e riuscì perfino a ferire l'imperturbabile Bohr con alcune osservazioni taglienti. Naturalmente per il venticinquenne Heisenberg la posta in gioco era molto alta: le sue nuove idee, i suoi pro-

L'esperimento ideale del microscopio a raggi gamma

Per illustrare il principio d'indeterminazione, Heisenberg ricorse a un esperimento ideale. Usando un microscopio di elevata risoluzione, che per l'illuminazione sfruttava i raggi gamma, egli tentò di dimostrare che la posizione e la quantità di moto dell'elettrone obbedivano al principio d'indeterminazione. Benché Heisenberg avesse ricavato il risultato corretto, Bohr osservò che nell'esperimento originale erano stati trascurati due punti essenziali: il potere di risoluzione del microscopio e il dualismo onda-particella. Nella versione corretta, un elettrone libero è situato direttamente sotto l'obiettivo del microscopio. L'elettrone vede la lente circolare sotto un cono di apertura 2Θ. L'elettrone viene poi illuminato da un raggio gamma proveniente da sinistra. Secondo i principi dell'ottica ondulatoria il microscopio può risolvere oggetti fino a una grandezza Δx, che è legata a Θ e alla lunghezza d'onda λ dalla formula:

$$\Delta x = \frac{\lambda}{2 \sin \Theta}.$$

Nell'istante in cui la luce viene diffratta nell'obiettivo del microscopio, l'elettrone si sposta a destra. Dopo l'urto, il raggio gamma osservato potrebbe essere diffuso secondo qualsiasi angolo entro il cono 2Θ. Nel caso estremo di diffusione verso il bordo anteriore (destro) della lente, la componente della quantità di moto nella direzione x sarebbe

$$p'_x + \frac{h}{\lambda'} \sin \Theta,$$

dove p'_x è la componente secondo x della quantità di moto dell'elettrone, λ' è la lunghezza d'onda del raggio gamma deviato, h è la costante di Planck (che lega la frequenza del fotone alla sua energia) e h/λ' è la quantità di moto totale del fotone del raggio gamma. Nell'altro caso limite, il raggio gamma diffonde all'indietro, colpendo il bordo sinistro della lente. In questo caso la componente secondo x della quantità di moto totale è

$$p''_x - \frac{h}{\lambda''} \sin \Theta.$$

In entrambi i casi la componente secondo x della quantità di moto finale dev'essere uguale alla stessa componente della quantità di moto iniziale.

Quindi:

$$p'_x + \frac{h}{\lambda'} \sin \Theta = p''_x - \frac{h}{\lambda''} \sin \Theta.$$

Se l'angolo Θ è piccolo, allora $\lambda' \sim \lambda'' \sim \lambda$:

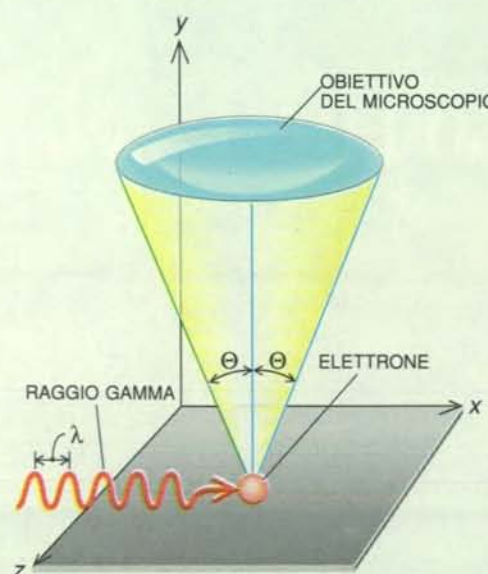
$$p''_x - p'_x = \Delta p_x \sim \frac{2h}{\lambda} \sin \Theta.$$

Poiché $\Delta x = \lambda/2 \sin \Theta$, vi è una relazione di reciprocità tra le indeterminazioni minime delle posizioni dell'elettrone lungo l'asse x e delle componenti della sua quantità di moto lungo lo stesso asse:

$$\Delta p_x \sim \frac{h}{\Delta x}.$$

Se l'incertezza è maggiore di questo minimo, si ricava la disuguaglianza

$$\Delta p_x \cdot \Delta x \geq h$$

che approssima la relazione d'indeterminazione di Heisenberg.

getti accademici e fors'anche il suo desiderio di essere pari ai suoi mentori sul piano intellettuale. In maggio il suo articolo comparve senza modifiche su un'importante rivista tedesca di fisica, ma conteneva un breve poscritto in cui si ammetteva l'errore del microscopio e si segnalavano ai lettori alcuni punti essenziali del ragionamento di Bohr.

All'età di 65 anni Heisenberg fu invitato a tenere alcune conferenze all'Università di Lipsia dove aveva a lungo svolto le proprie ricerche. Pochi anni dopo si ammalò e nel 1976 morì di cancro.

Circa quattro mesi dopo, Heisenberg si era asciugato le lacrime e il suo tono era completamente cambiato: adesso sembrava grato a Bohr per le sue critiche. Quando nel settembre 1927 Bohr illustrò per la prima volta la complementarità al Convegno internazionale di fisica di Como, Heisenberg, prima così sicuro dell'indeterminazione, gli espresse il primo dei suoi generosi riconoscimenti. Nella versione pubblicata della discussione che seguì l'esposizione di Bohr a Como, Heisenberg lo ringraziava per aver chiarito l'indeterminazione «in ogni particolare» e per aver enunciato quella che fu poi chiamata interpretazione di Copenaghen.

Può darsi che l'improvviso cambiamento nell'atteggiamento di Heisenberg dipendesse dal fatto di aver raggiunto i suoi traguardi. Nello stesso mese del Convegno di Como, aveva infatti saputo della sua imminente chiamata alla cattedra di Lipsia: quella meta era stata infine raggiunta.

Mentre la brama di mettere in luce le proprie capacità e i propri contributi alla meccanica quantistica si attenuava, in Heisenberg emergeva un altro desiderio, che ora riguardava anche Bohr: istituire a Lipsia un programma di ricerca di altissimo livello sui fondamenti della fisica. Oltre a consolidare il concetto ancora gracile di indeterminazione, le idee di Bohr fornivano un punto di riferimento ai seguaci della scuola danese che, al pari di Heisenberg, vagheggiavano una fisica compiuta da poter diffondere dalle loro cattedre appena acquisite e sfruttare nei loro articoli. Heisenberg e gli altri discepoli di Bohr non si mettevano più al servizio di singoli programmi e scoperte, come la meccanica delle matrici o l'indeterminazione, bensì dello «spirito di Copenaghen».

Heisenberg e gli altri riuscirono a far accettare la loro interpretazione nonostante le perduranti obiezioni di maestri come Einstein e Schrödinger. Nel lustro che seguì il Convegno di Como e la successiva Conferenza Solvay, Heisenberg e il gruppo del suo istituto formularono importanti teorie quantistiche sullo stato solido cristallino, sulla struttura molecolare, sulla diffusione della radiazione da parte dei nuclei e sulla struttura nucleare degli atomi. Con altri teorici, essi fecero enormi progressi verso una teoria quantistica relativistica dei campi e gettarono le fondamenta delle ricerche sulla fisica delle alte energie.

Questi risultati attrassero naturalmente molti tra i migliori studenti verso istituti come quello di Heisenberg. Questi studenti, nutriti della dottrina di Copenaghen, formarono una nuova generazione dominante di fisici che, quando dovettero emigrare in seguito all'ascesa al potere di Hitler negli anni trenta, portarono le loro idee in tutto il mondo.

Heisenberg e altri della scuola di Copenaghen non persero tempo nel portare la loro dottrina a quanti non visitavano gli istituti europei. In particolare Heisenberg trovò negli Stati Uniti un terreno fertile per fare proseliti. In un giro del mondo compiuto nel 1929 insieme a Dirac, Heisenberg tenne all'Università di Chicago alcune conferenze sull'interpretazione di Copenaghen che ebbero un effetto enorme. Nella prefazione della pubblicazione che fece seguito a queste conferenze, egli scrisse: «Questo libro avrà raggiunto il suo scopo se contribuirà in qualche misura alla diffusione di quel *Kopenhagener Geist der Quantentheorie*... che ha guidato tutto lo sviluppo della fisica atomica moderna.»

Di ritorno a Lipsia, l'araldo di questo *Geist* vide che le sue prime idee scientifiche erano adesso ampiamente accettate dalla comunità dei fisici, che gli accordò una posizione di rilievo sia sotto il profilo istituzionale sia sotto quello scientifico. Nel 1933 questa comunità conferì a Heisenberg, insieme con Schrödinger e Dirac, il riconoscimento più ambito: il premio Nobel.

Benché oggi venga celebrato a ragione come uno dei massimi fisici dei tempi moderni, Heisenberg fu anche criticato per molti comportamenti tenuti dopo l'ascesa al potere di Hitler. Pur non iscrivendosi mai al partito nazionalsocialista, Heisenberg ricoprì rilevanti posizioni accademiche e divenne portavoce della cultura tedesca nei territori occupati. Rifiutando ripetutamente proposte di trasferimento all'estero, guidò le principali ricerche sulla fissione dell'uranio eseguite per conto del Terzo Reich. Dopo la guerra fornì per la sua condotta varie spiegazioni, che offuscarono ancora di più la sua reputazione all'estero. Questa sconcertante combinazione di comportamenti discutibili e di brillanti ricerche rispecchia la difficile situazione degli scienziati e della scienza in un secolo turbolento e talora brutale. Per Heisenberg, leale suddito della Germania che aveva indagato tanto in profondità i misteri della natura, era difficile cogliere e condannare le tragiche aberrazioni della sua patria. Morì di tumore al rene e alla cistifellea nel 1976, nella sua casa di Monaco.

BIBLIOGRAFIA

FINE ARTHUR, *The Shaky Game: Einstein, Realism and the Quantum Theory*, University of Chicago Press, 1986.

MOORE WALTER J., *Schrödinger: Life and Thought*, Cambridge University Press, 1989.

CASSIDY DAVID C., *Uncertainty: The Life and Science of Werner Heisenberg*, W. H. Freeman and Company, 1991.

PAIS ABRAHAM, *Niels Bohr's Times: In Physics, Philosophy and Polity*, Oxford University Press, 1991.

ARCHEOLOGIA MOLECOLARE: UNA NUOVA DISCIPLINA

di Philip E. Ross

Gli acidi nucleici e le proteine, contenuti in mummie e in resti ossei risalenti a epoche remote, possono essere considerati veri e propri documenti storici in grado di svelare molti segreti del passato.

«I discendenti di questo poveretto vorranno senza dubbio sapere qual è stata la sua fine», pensava Helmut Simon scattando fotografie del cadavere che, l'estate scorsa, durante una vacanza in Tirolo con la moglie Erika, aveva visto affiorare dal ghiacciaio del Similaun. Simon aveva ragione: quei resti sono stati datati ed è stata loro plausibilmente attribuita un'età compresa tra i 5000 e i 5500 anni, il che rende quell'individuo un possibile antenato di milioni di europei viventi. Per questo motivo, gli studiosi si sono affannati per raggiungere i resti mortali e gli utensili di quella che è l'unica mummia nota dell'Età della pietra.

Malgrado il clamore sollevato attorno alla scoperta dell'uomo del Similaun, questo reperto costituisce soltanto un singolo tassello di un mosaico composto da migliaia di pezzi, che si sta lentamente costruendo. Incredibilmente ben conservati, talvolta associati a indumenti, armi e vettovaglie, con il contenuto del loro ultimo pasto spesso ancora ben riconoscibile nello stomaco, i corpi antichi si rinvengono con regolarità in certi siti propizi, dai deserti dell'Africa settentrionale, del Perù e degli Stati Uniti sudoccidentali fino alle torbiere della Danimarca, della Gran Bretagna e della Florida. Considerati per molto tempo alla stregua di semplici curiosità o di materia prima per il mortaio dello speziale, questi messaggeri mummificati del lontano passato sono stati sistematicamente distrutti o riseppelliti. Essendone stata compresa l'importanza, quelli che oggi rimangono sono sottoposti a un'indagine approfondita.

Il proverbiale silenzio di tomba è diventato un anacronismo poetico. Ben lungi dall'essere muti, gli antichi reperti offrono una testimonianza eloquente a coloro che sanno interrogarli. Nell'ultimo decennio, i biochimici hanno trovato il modo di estrarre le molecole della vita da residui organici appena tratti dalla polvere. Hanno recuperato proteine e acidi nucleici da piante vissute milioni di anni fa e da animali che risalgono a decine di migliaia di anni fa. Ma il fatto più incoraggiante di tutti è che attualmente gli scienziati cercano e trovano questi indizi molecolari non solo in corpi mantenutisi integri, ma anche in ossa fossilizzate, e ciò offre una panoramica ben più vasta sul passato.

Questa scienza giovane, che è l'archeologia molecolare, offre in assoluto la prima possibilità di verifica delle conclusioni a cui è giunta la genetica delle popolazioni moderne. Assieme agli studi comparati sulle lingue e sui manufatti, essa promette di dipanare la matassa intricata della storia biologica dell'umanità, delle divisioni e delle migrazioni, delle estinzioni e delle espansioni. L'archeologia molecolare ha fatto luce anche su ar-

Una mummia egizia predinastica del quarto millennio a.C. appare qui in una ricostruzione della tomba in cui è stata rinvenuta. Le mummie offrono la prima testimonianza diretta delle relazioni genetiche tra le popolazioni primitive.

gomenti d'interesse più pratico: i climi che le prime popolazioni dovevano affrontare, le malattie cui andavano soggette, i cibi di cui disponevano, le modalità di allevamento della prole. Infine, questa nuova disciplina permetterà ai biologi di ripercorrere passo passo la teoria di Darwin attraverso la storia. «Oggi possiamo sognare di cogliere l'evoluzione molecolare con le mani nel sacco» esultava il compianto Allan C. Wilson dell'Università della California a Berkeley, un pioniere in queste ricerche.

Nessuno sa quanto indietro nel tempo possano sperare di risalire i *detective* del DNA, ma vi sono buone ragioni per essere ottimisti. Una foglia ancora verde di magnolia ha conservato il proprio DNA per più di 17 milioni di anni e gli scienziati ipotizzano che residui di DNA di ominide riescano a sopravvivere in un osso perlomeno finché questo finisce per diventare il proprio calco mineralizzato. A quanto si sa, i più antichi resti umani analizzati risalgono solo a 8000 anni fa, ma questo primato potrebbe già essere stato battuto.

Nei tre anni trascorsi da quando alcuni ricercatori dell'Università di Oxford hanno recuperato per primi DNA da vecchie ossa, è circolata insistentemente la voce che vi fossero gruppi che ricercavano tracce analoghe nei resti di ominidi arcaici. Nessuno, però, ammetterebbe di utilizzare tecniche analitiche non collaudate che distruggerebbero reperti d'incalcolabile valore. «Molti studiosi hanno preferito rinunciare ad esaminare materiali rari, sebbene in realtà siano disponibili parecchi esemplari di Neandertal» sostiene Bryan Sykes, membro del gruppo di Oxford.

Tuttavia quelle voci non erano prive di fondamento. Erik Trinkaus, paleoantropologo presso l'Università del New Mexico, sostiene di aver fornito una piccola quantità di osso di Neandertal a biochimici del Los Alamos National Laboratory, i quali stanno oggi cercando di recuperare da esso il DNA. Il reperto osseo, un frammento di vertebra appartenente a uno scheletro trovato a Shanidar, in Iraq, risale a circa 50 000 anni fa. Trinkaus era disposto a rinunciare al campione, dato che la natura frammentaria di questo non gli permetteva neppure di identificare la vertebra di provenienza.

Fino alla fine di marzo, il gruppo di Los Alamos non avrebbe fatto riferimento alla ricerca, ma da fonti vicine al progetto si è appreso che, da studi preliminari, il DNA risultava presente, appariva danneggiato, forse per l'età, e proveniva da un primate. Le stesse fonti avvertivano però della necessità di effettuare le prove con estrema cautela, per escludere la possibilità che il DNA derivasse da contaminanti moderni e non già dall'osso antico. Soltanto allora il gruppo avrebbe finalmente pubblicato i propri risultati.

Che i ricercatori di Los Alamos siano stati o meno i primi a recuperare il DNA da resti umani antichi, il loro esperimento non è che un esempio del rapido progresso compiuto dall'archeologia molecolare. Fino a poco tempo fa, la maggior parte delle ricerche sul passato genetico dell'umanità si era concentrata non sulla pelle e sulle ossa dei corpi mummificati e degli scheletri, ma sul sangue delle popolazioni viventi. Non era disponibile alcuna valida alternativa per l'analisi di caratteri puramente ereditari. Innanzitutto, si sono studiate le proteine del sangue, specialmente gli antigeni polimorfi (come il fattore Rh) e le immunoglobuline. Poi, circa dieci anni fa, lo sviluppo di nuove tecniche biochimiche ha spostato l'attenzione dalle proteine al DNA nucleare che le codifica.

Nel 1984, Wilson e il suo gruppo a Berkeley hanno identificato per primi i geni nei vecchi tessuti, clonando il DNA estratto dalla cute del quagga, un animale simile alla zebra, estintosi in Africa un secolo fa. L'anno successivo, Svante Pääbo, un ricercatore svedese che attualmente insegna all'Università di Monaco di Baviera, riuscì a clonare il DNA di una mummia egizia risalente a oltre 4400 anni fa. Quindi nel 1988, mentre si trovava nel laboratorio di Wilson, applicò per primo la reazione a catena della polimerasi ad antichi resti umani. Amplificando il DNA mitocondriale estratto da un cervello conservato a Little Salt Spring, in Florida, si spinse più indietro di ben 2600 anni nei segreti del passato. Infine, nel 1989, Sykes e i suoi collaboratori Erika Hagelberg e Robert E. M. Hedges di Oxford hanno amplificato per primi il DNA estratto da ossa umane.

Nel 1985, Svante Pääbo dell'Università di Monaco di Baviera ha dimostrato l'attendibilità dell'archeologia molecolare, clonando il DNA di una mummia. Qui mostra un campione utilizzato nella sua attuale indagine genetica sulla storia egizia.

È stato pressoché impossibile ricavare tracce di DNA dagli antichi resti fino a quando, nel 1983, fu sviluppata la tecnica di reazione a catena della polimerasi, una «macchina copiatrice» per il DNA sorprendentemente sensibile (si vedano gli articoli *La scoperta della reazione a catena della polimerasi* di Kary B. Mullis in «Le Scienze» n. 262, giugno 1990 e *Gli acidi nucleici nei semi antichi* di Franco Rollo in «Le Scienze» n. 283, marzo 1992). Il processo si svolge in due tempi: in primo luogo, la doppia elica di un campione del DNA bersaglio viene scissa nei due filamenti singoli; quindi gli enzimi costruiscono un secondo nuovo filamento a partire da una soluzione di basi libere degli acidi nucleici. Attraverso la ripetizione di questo processo, un'unica molecola di DNA dà origine a due molecole di DNA, quindi a quattro, e via di seguito: un'amplificazione geometrica che può in teoria produrre una grande quantità di materiale per l'analisi. Ne consegue che i campioni di DNA che sarebbero sfuggiti a ogni identificazione una decina di anni fa danno origine oggi a bande soddisfacenti se analizzati con l'elettroforesi su gel, una tecnica basata sulla migrazione differenziale in un substrato di sostanze di peso molecolare diverso.

La grande sensibilità della reazione a catena della polimerasi rende questo mezzo di indagine quasi diabolico. Se anche solamente una cellula cutanea dello sperimentatore venisse lasciata cadere nella soluzione, la reazione a catena ne amplificherebbe il DNA e, a causa del miglior stato di conservazione di questo, lo farebbe in modo preferenziale. Dopo alcuni cicli di raddoppio, si scorgerebbero così nella provetta non i segreti dell'uomo di Neandertal, ma i propri segreti. «La maggior parte del DNA estratto da vecchi materiali è così degradata che quando mi dicono di aver ottenuto grandi bande su un gel, non ci credo» dice Rebecca L. Cann, ex allieva di Wilson e oggi all'Università di Hawaii, a Manoa.

Pääbo ha messo a punto metodi di laboratorio che riducono al minimo la possibilità di confondere il nuovo con il vecchio DNA. I ricercatori continuano a perfezionare queste tecniche da camera sterile, e intanto si esercitano in un'altra arte magica, cercando di avere partita vinta sugli inibitori, quelle sostanze che spesso impediscono alla reazione a catena della polimerasi di funzionare adeguatamente con il vecchio DNA. Per esempio, l'albumina del siero bovino riesce in qualche caso a sciogliere la sostanza viscosa che tiene arrotolato su se stesso il vecchio DNA.

Rimane comunque qualche motivo di frustrazione. Nel corso del primo convegno mondiale sul paleo-DNA, che si è svolto l'estate scorsa a Nottingham, in Inghilterra, gli addetti ai lavori hanno dovuto riconoscere che determinate analisi, che sono realizzabili sul DNA recente, non si possono realizzare su vecchi campioni di DNA. «Abbiamo cercato di stabilire il sesso degli individui ai quali apparteneva una serie di ossa sottoposte in precedenza allo stesso tipo di determinazione da parte di esperti di antropologia morfometrica» afferma Sykes. «Ma in una serie di prove in cieco di 20 scheletri non siamo mai riusciti ad ottenere il risultato giusto. Anche se, in realtà, non dovrebbe essere difficile.»

William W. Hauswirth dell'Università della Florida a Gainesville ha scoperto una sorprendente omogeneità genetica nel DNA di cervelli preistorici trovati in Florida.

I cervelli delle torbiere

Ed è già un compito tutt'altro che facile ritrovare campioni di DNA antico. Il DNA è stato ricavato dapprima da tessuti molli, che si sono conservati per essiccamento, come nel caso delle mummie egizie. Sono stati, invece, scartati i corpi che erano stati recuperati nelle torbiere, a dispetto del loro sorprendente grado di conservazione. Quegli uomini delle torbiere, così come le vittime sacrificali dell'Età del ferro, portate alla luce alcuni decenni fa in Danimarca, e un probabile druido, trovato sette anni fa in Gran Bretagna, mostrano squarci alla gola, tatuaggi sul corpo, cibo nell'apparato digerente e persino zampe di gallina attorno agli occhi. Purtroppo, il loro perfetto grado di conservazione è dovuto al fatto di aver subito l'azione dell'acido tannico, che distrugge il DNA.

L'uomo del Similaun, mummificato per congelamento, costituisce un'eccezione notevole ed estremamente insolita. I suoi resti non sono stati distrutti dal tannino; tutti gli indizi depongono a favore di una buona conservazione del DNA. L'ottimismo dei ricercatori si spinge fino all'auspicio di scoprire nei suoi tessuti batteri e altri organismi patogeni, forse addirittura spore vive (*si veda la finestra alle pagine 88 e 89*).

Occasionalmente, tuttavia, anche una torbiera può riservare qualche sorpresa. È questo il caso di Windover, in Florida, dove le sorgenti di acque calcaree, tamponando l'acido, hanno conservato il più antico DNA umano che mai sia stato campionato. Gli operai impegnati a scavare sul fondo della torbiera hanno trovato uno strato, dello spessore di 30 centimetri, corrispondente alla sedimentazione di un millennio, nel quale giacevano scheletri, alcuni dei quali con crani intatti e cervelli ben conservati, per quanto ovviamente raggrinziti. I più antichi risalivano a 8000 anni fa, i più recenti a 7000.

Nella prima metà degli anni ottanta, Glen H. Doran della Florida State University ha diretto un gruppo di antropologi impegnati nella realizzazione di un progetto ambizioso: scavare la torbiera, sezione per sezione, costruendo argini, mentre le pompe drenavano l'acqua. Sono stati così riportati alla luce alcuni scheletri ancora protetti da pali,

posti in croce sopra di loro con una certa angolazione forse per evitare che i corpi galleggiassero, facile preda di animali saprofagi.

Dopo aver recuperato 177 scheletri e 91 cervelli, gli archeologi della Florida State University hanno richiuso lo scavo, decidendo di lasciare intatta un'altra sessantina di esemplari. «A tutta prima, è stato difficile dire agli archeologi di comportarsi in quel modo» racconta William W. Hauswirth, microbiologo all'Università della Florida a Gainesville. «Ma sappiamo per certo che, *in situ*, il materiale genetico è stabile. Quando ho fatto rilevare che non sapevamo ancora come analizzare quel materiale - essi erano infatti convinti che sapessimo esattamente quello che stavamo facendo - sono stati d'accordo.»

Per quattro anni, Hauswirth dovette lottare contro le frustrazioni della clonazione vecchio stile e, quindi, contro le imprevedibili complicazioni della modernissima reazione a catena della polimerasi. Infine, un anno fa, in collaborazione con Cynthia D. Dickel, anch'essa a Gainesville, e con David A. Lawlor e Peter Parham della Stanford University, ha identificato segmenti di DNA particolarmente significativi, che erano stati prelevati da un uomo di Windover risalente a 7500 anni fa. I segmenti provengono dal maggior complesso di istocompatibilità (MHC, dall'inglese *Major Histocompatibility Complex*), che regola il sistema immunitario. L'MHC contiene molti siti, nei quali si trova o l'uno o l'altro gene di una coppia di geni alternativi, o alleli. «Possiamo caratterizzare gli alleli MHC presenti nelle antiche popolazioni e, alla fine, potremo essere in grado di dedurre a quali malattie esse fossero resistenti e a quali no» sostiene Hauswirth.

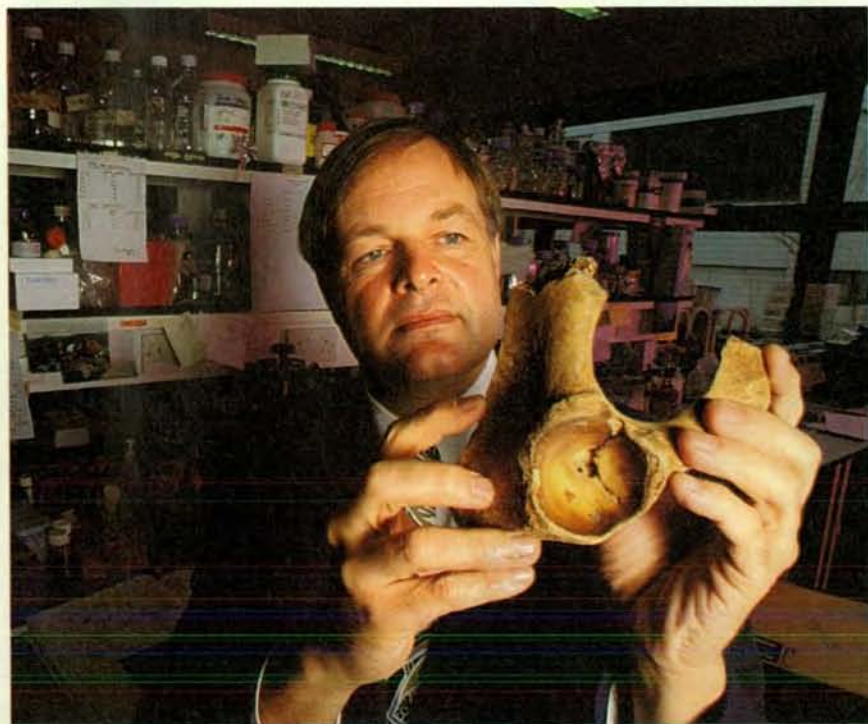
Il gruppo di Hauswirth ha trovato che la genetica della popolazione di Windover si è modificata moderatamente nei mille anni di storia testimoniati da quel luogo di sepoltura; pertanto in quella comunità devono essere stati contratti ben pochi matrimoni con individui che non ne facevano parte. «Stanno emergendo aspetti molto diversi da quelli delle popolazioni moderne» continua Hauswirth. «Abbiamo preso in esame tre loci genici e abbiamo trovato che sono molto più omogenei che nelle tribù contemporanee studiate.»

Questa omogeneità può voler dire che, nella popolazione di Windover, il grado di inincrocio era molto elevato. Se questo carattere fosse stato generale (cioè se i primi americani fossero rimasti tutti nella propria cerchia, sposandosi tra vicini), l'isolamento avrebbe potuto dare origine a una molteplicità di lingue, un quadro che oggi si può riscontrare nelle vallate delle regioni montuose della Nuova Guinea. Un simile scenario potrebbe spiegare perché, nel 1492, il Nuovo Mondo avesse una diversità linguistica altrettanto elevata di quella del Vecchio Mondo, così come potrebbe spiegare perché la Nuova Guinea ospiti ancora un quinto delle famiglie linguistiche di tutto il mondo.

Per verificare una simile argomentazione, gli studiosi possono confrontare la diversità genetica tra uomini e donne negli antichi luoghi di sepoltura. Secondo Lyle W. Konigsberg dell'Università del Tennessee, che sta studiando i costruttori di tumuli dell'Illinois, «all'interno di un singolo villaggio c'è da attendersi che, alla nascita, maschi e femmine presentino uguale variabilità genetica». E prosegue: «C'è anche da attendersi che ogni villaggio sia leggermente diverso dagli altri. Così, se per esempio le donne fossero provenute da altri luoghi, sarebbero state meno simili tra loro rispetto agli uomini». Questo è il quadro che Konigsberg ha notato, cercando l'esogamia tra i costruttori di tumuli.

D'altra parte, la tribù della Florida può semplicemente essere passata attraverso quello che viene definito un «collo di bottiglia» genetico. Una situazione del genere si determina quando una pestilenza, una carestia o una guerra decima una comunità o quando un ristretto gruppo si insedia in una colonia. In ciascuno di questi casi, la nuova popolazione sarà meno differenziata geneticamente rispetto alla popolazione d'origine e i geni saranno presenti in essa in rapporti differenti rispetto a prima. Il rimescolamento genetico produce quello che gli evoluzionisti chiamano «effetto del fondatore». Effetti di questo tipo, forse associati a una selezione naturale o sessuale, possono spiegare perché tanti irlandesi abbiano capelli rossi o perché gli abitanti delle Isole Samoa tendano a essere così robusti.

Parecchi gruppi di ricerca applicano oggi metodi genetici per trovare la prova del collo di bottiglia genetico nel quale gli asiatici dovettero imbattersi trasferendosi nel Nuovo Mondo. Pääbo e collaboratori dell'Università dello Utah stanno programmando un'esauriente indagine genetica sulle popolazioni viventi come pure sulle popolazioni antiche alle quali sono da attribuire le mummie degli Stati Uniti sudoccidentali risalenti a 900 anni fa. «Stiamo cercando di ricostruire la storia delle popolazioni dell'America Settentrionale e Meridionale» sostiene Pääbo. «Vogliamo vedere come queste popolazioni siano correlate l'una all'altra e come siano correlate a quelle asiatiche; vogliamo inoltre controllare se il loro albero genealogico coincida con le classificazioni linguistiche.»

Nel frattempo, Hauswirth, Noreen Tuross, una biochimica che lavora alla Smithsonian Institution, e altri ricercatori allestiranno tre banche dati per il DNA antico allo scopo di studiare gli effetti del morbillo e del vaiolo che uccisero la maggior parte degli indigeni americani nell'arco di poche generazioni in seguito al primo contatto con gli europei. I dati provengono da Windover, da Indian Knoll, nel Kentucky, che vanta 1200 scheletri ben conservati, e da Arroyo Seco, in Argentina. Quest'ultimo sito è sta-

Bryan Sykes e i suoi collaboratori dell'Università di Oxford hanno esteso il campo di applicazione dei dati paleogenetici, recuperando DNA da ossa essiccate.

Noreen Tuross della Smithsonian Institution di Washington congela le ossa e le macina, polverizzandole completamente, allo scopo di trovare antichi anticorpi.

to incluso sia per la ricchezza del materiale sia per la distanza dagli altri siti. Secondo Hauswirth, «questa sarebbe la prima analisi genetica dei contatti fra popolazioni indigene americane prima dell'avvento degli europei».

Indagini di questo tipo sul paleo-DNA dovrebbero confermare, o indebolire, le recenti teorie sull'uomo primitivo, basate su confronti genetici tra popolazioni moderne. Quando Luigi Luca Cavalli-Sforza e collaboratori, a Stanford, hanno rappresentato i rapporti genetici a livello di popolazione sotto forma di alberi genealogici, è risultato che questi alberi si correlavano bene con una classificazione genetica delle lingue (si veda l'articolo *Geni, popolazioni e lingue* di Luigi Luca Cavalli-Sforza in «Le Scienze» n. 281, gennaio 1992). Il loro lavoro si è concentrato sulla preistoria recente, cioè a partire dal periodo successivo alla comparsa di *Homo sapiens* di tipo moderno.

L'estraneità del padre

Fatto della massima importanza, lo studio del paleo-DNA promette di verificare la teoria di Wilson e collaboratori in base alla quale tutti gli esseri umani sarebbero discesi da un'unica donna, Eva, che, secondo le loro conclusioni, sarebbe vissuta in Africa. Wilson, la Cann e Mark Stoneking hanno analizzato il DNA dei mitocondri, organelli cellulari che trasformano il glucosio in una forma di energia meglio utilizzabile. Diversamente dal DNA del nucleo cellulare, il DNA dei mitocondri viene trasmesso inalterato di madre in figlio. (Anche se studi effettuati l'anno scorso sui topi hanno dimostrato che singoli mitocondri paterni possono passare nella generazione futura, questi non mescoleranno mai i propri geni con quelli materni.) L'estraneità del padre rende le linee di discendenza ben distinte, cosicché diviene possibile ricostruire la genealogia semplicemente come risultato di mutazioni.

Wilson e collaboratori hanno confrontato il DNA mitocondriale di popolazioni di tutto il mondo per stimare le distanze genetiche. Hanno poi indicato queste distanze in un albero genealogico, al quale hanno messo le radici confrontando i dati con quelli provenienti dagli scimpanzé. Dato che i rami più lunghi erano in Africa (il che suggerirebbe che il DNA mitocondriale ha cominciato a differenziarsi precisamente in quel continente), è proprio lì che essi hanno collocato Eva. Anche se nessuno mette in discussione la preminente diversità dei geni africani, recentemente alcuni studiosi hanno trovato da ridire sull'albero. Essi hanno dimostrato che metodi statistici alternativi permettono di ricostruire alberi genealogici dell'uomo ugualmente plausibili con radici in Asia.

Fatto ancora più controverso, il gruppo di Wilson ha desunto un «orologio molecolare» dalle differenze genetiche accumulate da esseri umani e scimpanzé nei loro cinque milioni, o poco più, di anni di vita separata. Il DNA mitocondriale si presta a una simile cronometria perché la sua rapida velocità di mutazione ne fa un indice temporale preciso. L'orologio ha fatto risalire Eva a circa 200 000 anni fa, inducendo il gruppo di Wilson a concludere che i suoi discendenti (i primi uomini moderni) si fossero diffusi dall'Africa per sostituire altri ominidi con scarso o addirittura nullo esoincrocio (si veda l'articolo *Una genesi africana recente* di Allan C. Wilson e Rebecca Cann in «Le Scienze» n. 286, giugno 1992).

La teoria della «diffusione dall'Africa» era stata formulata in maniera del tutto indipendente, su testimonianze puramente paleontologiche, da Christopher B. Stringer del Natural History Museum di Londra (si veda l'articolo *La comparsa dell'uomo moderno* di Christopher B. Stringer in «Le Scienze» n. 270, febbraio 1991). Da allora essa è stata soggetta alle critiche di altri studiosi, che interpretano gli stessi fossili in maniera diversa (si veda l'articolo *Un'evoluzione multiregionale* di Alan G. Thorne e Milford H. Wolpoff in «Le Scienze» n. 286, giugno 1992).

Per finire, gli esperti di genetica molecolare sperano di risolvere il dibattito con analisi del DNA conservatosi in antiche ossa. Le banche dati genetiche, come quelle proposte da Hauswirth, potrebbero assicurare la taratura dell'orologio mitocondriale in termini assoluti. Si dovrebbero trovare i discendenti moderni del paleo-DNA, conteggiare le sostituzioni di coppie di basi e confrontarle con le datazioni al radiocarbonio delle ossa in cui sia stato trovato il DNA.

Come punto di partenza, Pääbo e Anna Di Rienzo, sua collega nel laboratorio di Berkeley, studieranno sia mummie egizie sia le popolazioni moderne che vivono in località lungo il Nilo, e que-

sto per cercare indizi di una migrazione dall'Africa. Secondo la Di Rienzo, «il Nilo avrebbe facilitato lo spostamento delle popolazioni». Essa aggiunge: «Sono certa che quando disporremo di maggiori informazioni, troveremo un gradiente» da sud a nord. (Una verifica ancora più significativa potrebbe venire da un unico campione di DNA di Neandertal. Se, per esempio, questo campione avesse marcatori che si trovano solo in alcuni rami dell'albero genealogico di Eva, il risultato confuterebbe la teoria della sostituzione totale.)

Altri studiosi di archeologia molecolare stanno oggi lavorando su antichi reperti per ripercorrere gli effetti delle malattie sulle popolazioni umane. Noreen Tuross della Smithsonian Institution ricerca prove molecolari dirette di casi precolombiani di treponematosi, infezioni provocate dalle spirochete, i batteri a forma di elica che provocano la sifilide venerea, la sifilide endemica, la framboesia e la pinta. (La Tuross ammette comunque che i suoi test non permettono in teoria di distinguere un ceppo di treponemi dall'altro.)

Invece di andare a caccia del DNA, la Tuross cerca anticorpi nelle ossa di individui gravemente infettati. Finora ha identificato le immunoglobuline della classe IgG in due gruppi di scheletri del Nuovo Mondo: quello di Windover e un altro ritrovato nelle Great Plains settentrionali, in un sito del XVI secolo. Alla fine, uno studio molecolare comparato su infezioni da treponemi nel Vecchio e nel Nuovo Mondo potrebbe risolvere la questione se la sifilide sia stata davvero portata in Europa dall'equipaggio di Colombo.

Tom D. Dillehay, un archeologo dell'Università del Kentucky, sta cercando di ricostruire i colli di bottiglia dovuti a malattie, nei quali si sono imbattuti i primi popoli che sono arrivati nel Nuovo Mondo. Egli ritiene che la migrazione verso le regioni meridionali attraverso il Canada, in un clima rigido, possa aver liberato le popolazioni da alcuni parassiti del Vecchio Mondo. Tuttavia, c'erano sicuramente nuovi parassiti pronti ad assalirli, specialmente nei climi umidi del Messico e dell'America Centrale. Dillehay è convinto dell'importanza che, sotto il profilo archeologico, può avere l'ecologia delle malattie, ma sostiene che non è stato affatto semplice comprenderlo.

Tra i colli di bottiglia più interessanti, dovuti ad agenti infettivi, si annoverano quelli che selezionano le malattie genetiche. Una famiglia di malattie del sangue, le talassemie, è comparsa sotto forma di mutazioni indipendenti in Asia e nell'area del Pacifico. Siti preistorici in Thailandia hanno appena fornito agli storici della talassemia la più antica prova fondata su una base «solida», l'osso. «Riteniamo di avere la prova che 4000 anni fa la popolazione era già colpita da

Un messaggero dell'Età della pietra

Quando, da entrambe le parti del confine tra Austria e Italia, le polizie ebbero per la prima volta la notizia che un cadavere era stato scoperto nel ghiacciaio alpino del Similaun, reagirono con una certa indifferenza. In fin dei conti, i ghiacciai restituiscono periodicamente i loro morti; nel solo 1991, il clima insolitamente caldo aveva fatto scongelare una decina di cadaveri. Il tempo volgeva al brutto e il fine settimana era vicino. Perché precipitarsi a recuperare un morto?

Un paio di ottimi scalatori, incuriositi da quella prima notizia, pensarono, invece, di indagare per conto proprio e riportarono notizie sorprendenti: il corpo, per metà a mollo in una pozza d'acqua, presentava sul dorso quattro linee parallele bluastre. Ai lati del corpo giacevano strani oggetti, tra cui una scarpa e un arco senza corda. Era opportuno ora darsi un poco da fare. Rainer Henn dell'Istituto di medicina legale di Innsbruck si inerpicò fino a raggiungere il ghiacciaio, riconobbe immediatamente che quel cadavere era una mummia e lo compose in modo da poterlo trasportare con un elicottero fino al suo laboratorio.

Fu un primo punto a vantaggio dell'Austria. Austria e Italia avrebbero poi disputato per mesi in merito ai diritti sull'uomo del Similaun. Gli austriaci dirigono oggi i due gruppi internazionali che si sono spartiti lo studio dell'uomo del Similaun e dei suoi oggetti. Werner Platzer, anatomo-patologo dell'Università di Innsbruck, dirige il gruppo che analizza la mummia. Konrad Spindler, archeologo di Innsbruck, dirige il gruppo che studia i segni presenti sul corpo e l'equipaggiamento.

L'uomo del Similaun portava ancora una collana con un pendente di pietra bianca. «Questo piccolo pezzo di pietra non ha alcuna funzione e deve perciò avere un significato suo proprio», osserva Spindler. «Ritengo che sia un amuleto.» I segni sulla pelle che, secondo lui, sarebbero tatuaggi colorati con carbone, «sono i più antichi tatuaggi conosciuti. Quelli immediatamente meno vecchi, nella graduatoria, avrebbero solo 2000 anni circa». Possono significare qualunque cosa, fa notare Spindler: appartenenza a una certa famiglia, tribù, villaggio o ceto.

Per quanto riguarda l'ultimo punto, almeno, vi sono indizi che l'uomo del Similaun fosse benestante. Indossava abiti in pelle di buon taglio, foderati di paglia per difenderlo dai rigori delle alte quote. Possedeva, inoltre, un'ascia in metallo, un oggetto anch'esso di pregio. All'inizio gli esperti pensavano che si trattasse di un'ascia di bronzo e che pertanto l'uomo del Similaun risalisse a circa 4000 anni fa. Le analisi hanno, invece, dimostrato che l'ascia è di rame, metallo tipico del tardo Neolitico.

Questo indizio cronologico ha trovato conferma nella datazione degli indumenti con carbonio radioattivo, che ha fatto arretrare l'età dell'uomo del Similaun a circa 5000 anni. Finalmente, in febbraio, la datazione della cute e delle ossa ha fissato l'età del fossile ancora alcune centinaia di anni più indietro.

Si ha una scarsa esperienza nello studio del DNA in esseri umani disidratati per congelamento, ma secondo gli esperti il DNA dovrebbe conservarsi almeno altrettanto bene che nelle mummie egizie e nei cervelli trovati nelle paludi della Florida. Inoltre il corpo dell'uomo del Similaun potrebbe conservare organismi parassiti, dei quali si potrebbe ricostruire la sequenza del DNA. Alcuni sperano persino di trovare spore quiescenti.

I primi risultati sull'uomo del Similaun avrebbero dovuto essere pronti entro aprile di quest'anno, anche se la ricerca di possibili legami genealogici con le popolazioni moderne richiederà sicuramente tempi molto più lunghi. Per ora, dice Spindler, si sa solo che, come i nove decimi dell'umanità attuale, quell'uomo aveva occhi neri e capelli scuri. (I capelli, caduti dopo la morte, misuravano poco meno di nove centimetri; abbastanza lunghi da scendere sugli occhi.)

I ricercatori stanno avanzando ipotesi ben fondate su alcuni degli enigmi posti dall'uomo del Similaun.

• **Che cosa faceva sulle montagne?** Avrebbe potuto cacciare con l'arco, dato che la faretra conteneva alcune frecce pronte per essere usate. Forse stava ricercando del rame. Forse stava trasferendo pecore e capre dalle praterie alpine, dove avviene il pascolo estivo. «Sappiamo che è morto alla fine dell'estate - dice Spindler - perché gli abbiamo trovato addosso una piccola prugna, del tipo che matura in settembre.»

• **Come morì?** Non aveva escoriazioni o ferite e le scintigrafie non mettono in luce alcuna anomalia nelle ossa o negli organi. I denti sono consumati, ma questo riflette il tipo di dieta: pane contenente sabbia abrasiva proveniente dalle grossolane pietre da macina usate allora. Si pensa che quell'uomo avesse un'età compresa tra i 20 e i 30 anni.

Infine, l'uomo del Similaun aveva con sé una buona scorta di cibo, tra cui un poco di carne. Non doveva certamente avere fame e pertanto deve essere morto per esposizione a condizioni ambientali avverse. «Penso che sia morto di notte, per il freddo», asserisce Spindler. «È molto pericoloso in queste montagne a questa altitudine. Se siete stanco e vi sdraiate per dormire, bastano poche ore per morire. Ogni anno succede qualche incidente di questo tipo.»

• **Come ha resistito il suo corpo?** Un improvviso temporale potrebbe averlo costretto a cercare riparo nella piccola depressione rocciosa dove è stato trovato. Il freddo lo ha ucciso e ne ha conservato la carne, prosciugata poi dal sole e dal vento. Poi una nevicata propizia potrebbe aver nascosto il corpo alla vista degli uccelli, che hanno dovuto accontentarsi solo di qualche beccata. Infine, neve e ghiaccio si sono depositati, formando uno strato di parecchi metri, non sufficientemente spesso, tuttavia, da schiacciare quel corpo indurito, il quale è rimasto dove era perché la depressione rocciosa lo ha protetto, con la neve circostante, dai ghiacci che al disopra esercitavano lentamente la loro azione abrasiva.

Gli archeologi riprenderanno a scavare nella zona, ma non si pensa di trovare compagni dell'uomo del Similaun, ammesso che ne avesse, accomunati con lui nella morte. Egli ci potrebbe dire, con le parole del servo di Giobbe: «Sono scampato io solo, che ti racconto questo.»

Nel suo sepolcro di ghiaccio, l'uomo del Similaun appare come lo videro i primi turisti, con la testa e le spalle che sporgevano dal ghiacciaio. I biologi hanno controllato che il suo corpo non fosse contaminato da funghi (*a sinistra, in alto*), quindi lo hanno posto in condizioni analoghe a quelle che lo avevano preservato per oltre 5000 anni. L'usura dei denti (*a sinistra, in basso*) riflette una dieta a base di pane contenente sabbia, probabilmente proveniente dalle pietre da macina usate per ridurre in farina il grano. Finora i tatuaggi blu sulla schiena (*a destra, in alto*) si sono sottratti a qualsiasi interpretazione. L'abbigliamento e gli oggetti (*a destra, in basso*), giunti fino a noi senza il «filtro» dei riti funerari, danno una visione della vita quotidiana nell'Età della pietra.

talassemia», afferma Philip Houghton, un chirurgo passato all'antropologia, che lavora all'Università di Otago, in Nuova Zelanda. «Nelle ossa dei bambini in tenera età e dei giovani si notano segni di crescita anormale derivante da una iperproduzione talassemica di cellule ematiche.»

Collocando i processi evolutivi in un contesto temporale, questi risultati aiutano i ricercatori a perfezionare i propri modelli per verificare se altri geni siano comparsi per caso o attraverso una selezione. È stata così spiegata persino l'incapacità di molti adulti di digerire il lattosio, lo zucchero presente nel latte fresco. Anche se non è stata finora riscontrata alcuna prova molecolare, gli studiosi di storia biologica ritengono di poter dedurre dai dati contemporanei andamenti selettivi.

La lattasi, l'enzima che digerisce il lattosio, si conserva generalmente nella vita adulta in membri di gruppi etnici discendenti da società in cui venivano praticate attività lattiero-casearie. È generalmente assente, invece, nei membri adulti di gruppi che traggono sostentamento dalla caccia e dalla raccolta di cibo. Un'eccezione alla regola ha attirato l'attenzione degli studiosi: tra i raccoglitori di lingua san (più noti come boscimani), che vivono nel deserto africano del Kalahari, circa il 10 per cento degli adulti è in grado di digerire il lattosio. «I san, che bevono latte fresco, non avrebbero mai potuto essere genuini cacciatori-raccoglitori», ha concluso Michael J. Casimir dell'Università di Colonia in un articolo pubblicato nel 1990 sulla rivista «Current Anthropology».

Secondo Casimir, i raccoglitori si sarebbero staccati dai vicini ottentotti, rinunciando al costume di vita pastorale che questa popolazione segue ancora ai giorni nostri. Ma G. Eichinger Ferro-Luzzi di Roma avanza alcune obiezioni all'argomentazione di Casimir, rilevando: «È assolutamente improbabile che, per esempio, una popolazione come quella danese, la quale ha una delle più basse percentuali nel mondo di cattivo assorbimento del lattosio, possa mai aver subito uno stress dietetico tale per cui il bere o il non bere latte fresco potesse essere critico per la sua sopravvivenza».

Principali rinvenimenti di mummie umane

Il latte, ovviamente, è la bevanda dei neonati e dei bambini in tenera età e potrebbe avere sorprendenti implicazioni nel momento in cui essi vengono svezzati. Nelle moderne culture di raccoglitori, le madri allattano tipicamente la prole per diversi anni, durante i quali la lattazione sopprime la fecondità. Ma Jane E. Buikstra dell'Università di Chicago ha avanzato l'ipotesi che i primi agricoltori accelerassero lo svezzamento integrando l'allattamento al seno con una pappa a base di cereali. Questo tipo di alimentazione del bambino avrebbe permesso alle madri di cessare precocemente l'allattamento e quindi di essere in grado anticipatamente di procreare di nuovo, permettendo così alle società di agricoltori di sovrastare numericamente i raccoglitori che vivevano nelle loro vicinanze.

Nell'arco di alcuni millenni, quindi, gli agricoltori avrebbero sopraffatto i raccoglitori, mettendo sotto l'aratro i loro terreni di caccia e sostituendo i loro linguaggi con i propri, che andavano rapidamente differenziandosi. Avrebbero così disseminato vaste fasce di territorio con i germi delle famiglie di lingue indoeuropee, sino-tibetane, austronesiane e africane. Questa teoria è sostenuta da Colin Renfrew dell'Università di Cambridge (per quanto concerne le lingue indoeuropee) e da Peter Bellwood dell'Australian National University (per le lingue austronesiane).

Perché spiegare in questo modo la diffusione degli agricoltori? Perché altre argomentazioni non sembrano più valide di questa. Molti ricercatori concordano con Jared Diamond, un fisiologo dell'Università della California a Los Angeles, secondo il quale l'agricoltura potrebbe essere stata «il peggior errore nella storia della specie umana». Questi studiosi vedono l'agricoltura come un espediente di seconda scelta adottato per supplire a quel ricco approvvigionamento di cibo che la caccia non riusciva più a garantire, forse per lo sfruttamento intensivo dei branchi di animali delle praterie. Invece di disperdersi, la gente si legava alla terra. Diamond cita prove sempre più numerose, secondo cui le diete dei primi agricoltori, basate sulla monocoltura, così come il loro modo di vivere in comunità affollate, avrebbero reso i fisici più scarni, i denti più frequentemente soggetti a carie e la vita media più breve di quella dei raccoglitori.

Nel suo recente libro, *The Third Chimpanzee*, Diamond scrive che, quando, circa 6000 anni fa, l'agricoltura raggiunse quelle che sono oggi la Grecia e la Turchia, la statura degli uomini si abbassò di 17 centimetri e mezzo, raggiungendo una media di 157,5 centimetri, mentre quella delle donne si abbassò di 12 centimetri e mezzo, raggiungendo una media di 152,5 centimetri. «In epoca classica - egli aggiunge - le stature stavano di nuovo aumentando molto lentamente, ma greci e turchi moderni non hanno ancora riconquistato i livelli di statura dei loro robusti antenati cacciatori-raccoglitori.»

Pertanto, l'età alla quale avviene lo svezzamento della prole potrebbe essere un modo per documentare il passaggio da una società di raccoglitori a una società agricola. Si tratta di trovare marcatori chimici che permettano di distinguere i bambini ancora allattati al seno da quelli che già sono stati svezzati. Noreen Tuross e Marilyn L. Fogel della Carnegie Institution di Washington hanno messo a punto una tecnica per studiare non la matrice minerale dell'osso, ma le fibre di collageno che la percorrono, consolidando l'osso proprio come la paglia dà consistenza ai mattoni di argilla. Il collageno, come tutte le proteine sintetizzate dagli animali, concentra N-15, l'isotopo più pesante dell'azoto. Gli animali che si nutrono di altri animali concentrano ulteriormente questo isotopo. «Il lattante - spiega la Tuross - è un carnivoro, nel senso che "mangia la mamma". Così, sotto il profilo isotopico, si distingue da essa, mentre quando viene svezzato, diventa conforme a essa.»

Sono stati esaminati due campioni archeologici: l'uno che proviene da un sito preagricolo del Tennessee, risalente a un periodo compreso tra 6000 e 5000 anni fa, e l'altro proveniente da un sito del South Dakota, in cui dal 1650 al 1733 era stato coltivato mais. Un antropologo ha calcolato l'età dei corpi dei bambini trovati in ciascun sito in base allo sviluppo dei denti. La Tuross, invece, ne ha sciolto le costole, ottenendo una poltiglia gelatinosa contenente collageno.

Secondo questa analisi, l'introduzione della coltivazione del granoturco nelle Great Plains non ha avuto influenza sui tempi di svezzamento. L'analisi isotopica del collageno ha messo in rilievo che l'azoto pesante raggiungeva un massimo di concentrazione all'età di 1,1 anni, sia nei raccoglitori sia negli agricoltori, prima di scendere ai livelli dell'adulto nei successivi due anni di vita. La fase ulteriore, secondo la Tuross, consiste nell'esaminare esemplari provenienti dai principali centri di innovazione agricola: l'America Meridionale, la Mezzaluna Fertile, la Cina e la Nuova Guinea. Se una qualsiasi di queste regioni fornisse la prova che l'età dello svezzamento si era andata abbassando con la comparsa dell'agricoltura, l'ipotesi della Buikstra sarebbe confermata.

Anche la Polinesia è diventata un centro di interesse per i genetisti, in parte perché le sue lingue, che sono derivate dalla famiglia austronesiana, sembrano riflettere in modo straordinario migrazioni preistoriche. Questa caratteristica rende relativamente semplice la ricerca di marcatori genetici appropriati. Inoltre, i genetisti apprezzano l'incontaminata purezza dei geni polinesiani, specialmente di quelli incorporati nel DNA mitocondriale. Per Sykes, «essendo essi ereditati per via materna, non risento-

I geni di Neandertal potranno essere il test per Eva?

Nei nove anni trascorsi da quando i biochimici hanno cominciato ad amplificare tracce estremamente piccole di DNA in quantità che i genetisti riescono a studiare, l'orizzonte dell'archeologia molecolare si è spinto sempre più indietro nel passato. Molti progetti ora in corso cercano di studiare materiali che provengono dalla preistoria recente; sembra che altri tentativi più sommessi raggiungeranno nello spazio di mesi i primi successi, rivelando i genomi di esseri umani arcaici. La disponibilità di DNA antico, possibilmente risalente addirittura ai neandertaliani, permetterà agli antropologi di verificare numerose teorie sulle origini dell'umanità, in modo più diretto di quanto un tempo si ritenesse possibile.

Forse la più importante di queste teorie ricorre alla genealogia per determinare dove e quando siano comparsi gli esseri umani di aspetto moderno. Un albero genealogico, proposto nel 1987 da un gruppo guidato dal compianto Allan C. Wilson dell'Università della California a Berkeley, fa risalire il DNA mitocondriale, lungo la linea di discendenza materna, fino all'unico antenato comune di tutti gli esseri umani viventi.

Dato che i rami più profondi dividevano gli africani dai non africani e la diversità genetica era massima in Africa, il gruppo di Wilson ha fatto risalire anche Eva, come essa viene chiamata, all'Africa. Quindi, ricorrendo come indice cronologico a un confronto uomo-scimpanzé, ha datato Eva tra i 150 000 e i 200 000 anni fa, parecchio tempo dopo che gli ominidi arcaici si erano insediati nella maggior parte del Vecchio Mondo. I risultati hanno suggerito che i primi esseri umani moderni in Africa avevano sostituito altri ominidi senza incrociarsi con loro.

Alcuni paleontologi hanno accolto con entusiasmo la teoria; molti altri hanno fatto una dura opposizione. La critica di gran lunga più grave, tuttavia, è stata mossa dai genetisti, alcuni dei quali sostengono oggi che il gruppo di Wilson non è riuscito a rendersi conto che il programma di simulazione al calcolatore di cui si è servito per costruire l'albero può produrre migliaia, se non addirittura milioni, di alberi ugualmente plausibili, molti dei quali hanno le loro radici in Asia. «È probabile che l'ipotesi di Eva non sarebbe mai stata avanzata se, sull'insieme dei dati originali, fosse stata eseguita un'opportuna analisi filogenetica» dichiara Alan R. Templeton della Washington University.

Mark Stoneking, un membro del gruppo di Wilson, che insegna oggi alla Pennsylvania State University, afferma che la maggior diversità genetica degli africani continua ad essere un sostegno per l'ipotesi di Eva. Ma ammette che le nuove argomentazioni statistiche tolgono di mezzo il motivo principale su cui si basava l'argomentazione originale. «Dobbiamo staccarci dall'analisi degli alberi», egli dice. Una alternativa, aggiunge, sarebbe quella di passare in rassegna i dati per trovare le tracce di un improvviso aumento demografico, come quello che si sarebbe verificato in seguito alla grande diffusione dall'Africa.

«Si possono anche trovare i geni che controllano la variazione morfologica» aggiunge Stoneking. Per esempio, i geni responsabili dei caratteri fisici che differenziano le sottopopolazioni umane moderne, o razze. «Si possono poi confrontare questi geni del DNA antico con quelli del DNA moderno.» Se geni di questo tipo venissero trovati (e dimostrassero che gli ominidi arcaici d'Europa, Asia e Africa non assomigliano agli abitanti moderni di quelle rispettive regioni), la teoria delle recenti origini africane troverebbe un sostegno.

Questo neandertaliano di Shanidar, in Iraq, risale a circa 50 000 anni fa. Un frammento di osso di uno scheletro vicino è stato consegnato ai biochimici per cercare di recuperarne il DNA.

Anche un minimo successo nel recuperare geni da neandertaliani e da ominidi arcaici si alimenterebbe da sé, inducendo i paleontologi a mettere un maggior numero di ossa a disposizione dei biochimici. La risultante valanga di dati paleogenetici potrebbe immediatamente permettere ai ricercatori di costruire alberi genealogici di gran lunga migliori di quelli che si possono ricavare partendo dal DNA di esseri umani viventi.

Inoltre, l'informazione procederebbe oltre quella fornita anche dalla più perfetta genealogia, in quanto suggerirebbe non solo che cosa è accaduto, ma perché. I ricercatori, muniti degli antichi dati molecolari, potrebbero sperare di stimare i vantaggi selettivi conferiti da geni, o complessi di geni, particolari. Potrebbe anche essere soppesato il ruolo che la cieca casualità ha svolto nella sopravvivenza o nell'estinzione di genealogie. Nel suo complesso, la nuova ricerca darà sicuramente agli antropologi una visione più chiara dell'evoluzione del corpo, del cervello e forse anche della mente dell'uomo.

MUSICA E STRUMENTI MUSICALI

LE SCIENZE *edizione italiana di* SCIENTIFIC AMERICAN

ha dedicato all'argomento diversi articoli:

La fisica degli ottoni
di A. H. Benade
(n. 63, novembre 1973)

La dinamica musicale
di B.R. Patterson
(n. 78, febbraio 1975)

Illusioni musicali
di D. Deutsch
(n. 96, agosto 1976)

Le corde accoppiate del pianoforte
di G. Weinreich
(n. 127, marzo 1979)

L'acustica dei piani armonici di violino
di C. Maley Hutchins
(n. 160, dicembre 1981)

La fisica dei timpani
di T. D. Rossing
(n. 173, gennaio 1983)

La fisica delle canne d'organo
di N. H. Fletcher e S. Thwaites
(n. 175, marzo 1983)

La tromba barocca
di D. Smithers, K. Wogram e J. Bowsher
(n. 214, giugno 1986)

Il calcolatore e la musica
di P. Boulez e A. Gerzso
(n. 238, giugno 1988)

L'acustica del clavicembalo
di E. L. Kottick, K. D. Marshall e T. J. Hendrickson
(n. 272, aprile 1991)

no dell'immigrazione europea e cinese, nella quale furono coinvolti solo gli uomini».

«L'antico DNA mitocondriale corrisponde a quello di uno di due gruppi viventi principali che abbiamo trovato in Polinesia», sostiene la Cann. Analoghi profondi collegamenti sono stati documentati da Sykes tra gli abitanti delle Isole Cook e da Stoneking tra gli abitanti degli altopiani della Nuova Guinea.

I ricercatori hanno anche rilevato che i geni polinesiani sono meno variabili di quelli di altre popolazioni del Pacifico, il che fa pensare che gli antenati dell'attuale popolazione polinesiana siano passati attraverso un collo di bottiglia recente, «probabilmente perché solo gli occupanti di una singola barca, o di poche barche, sono riusciti a insediarsi», dice Stoneking, che oggi insegna alla Pennsylvania State University. Egli cita, tra gli altri dati, la «delezione di nove coppie di basi» nel DNA mitocondriale, una mutazione che ha avuto origine in Asia e che, in una certa misura, si osserva nella maggior parte delle popolazioni che provengono da questo continente. Questa caratteristica è particolarmente frequente sulle coste della Nuova Guinea e, soprattutto, in Polinesia.

Da prove come queste gli esperti di genetica molecolare deducono che gli emigranti provenienti dalle isole dell'Asia sudorientale dovettero scambiarsi i partner sessuali con le popolazioni delle coste della Nuova Guinea e che alcune donne proseguirono in seguito per la Polinesia. La Cann vuole verificare direttamente la teoria, ricostruendo la sequenza nucleotidica del DNA estratto da resti di scheletri in tutta la regione. Almeno in Polinesia questi resti erano spesso ben conservati grazie alla pratica di seppellire i morti nella sabbia asciutta o in cilindri di lava fredda e asciutta delle isole vulcaniche.

Parlando in nome dei morti

Rimane, tuttavia, l'imbarazzante problema dei diritti di proprietà: i morti appartengono allo scienziato che li trova o alla popolazione vivente che da quei progenitori discende? In alcune parti della Polinesia, i materiali per gli studiosi di archeologia molecolare abbonderebbero, fa notare mestamente la Cann, se soltanto gli scienziati potessero ottenere l'accesso a queste fonti. «Proprio ora disponiamo di buoni dati riguardanti uno scheletro trovato ad Hawaii, ma non possiamo pubblicarli», e questo per l'opposizione alla ricerca da parte degli indigeni hawaiiani.

«L'università si è così sensibilizzata al problema che agli osteologi che svolgono questo tipo di ricerca non viene neppure permesso di disporre del materiale nell'ambito dei loro laboratori.» Analoghe difficoltà affliggono gli archeologi australiani, che hanno soltanto potuto vedere gli scheletri di Kow Swamp tornare sotto terra. Si trattava dei più antichi resti umani datati su quel continente.

Gli studiosi di archeologia molecolare obiettano che quanto essi vanno cercando coincide esattamente con gli interessi degli attivisti dei diritti degli indigeni: una ricostruzione del progenitore e delle sue realizzazioni. A questo essi aggiungono i possibili benefici di ordine sanitario che deriverebbero dalla ricerca sulle origini delle predisposizioni genetiche alle malattie.

Ma, per molte popolazioni indigene, i benefici della scienza tendono a scomparire quando sono posti a confronto con le testimonianze talvolta mostruose sull'operato degli «scienziati» colonialisti. Quando, nel 1869, morì l'ultimo indigeno maschio della Tasmania, gruppi rivali di medici si disputarono pezzi di quel corpo. Un medico illustre utilizzò un frammento di pelle per farne un contenitore per il tabacco. Queste imprese ignobili fecero inorridire l'ultima donna tasmaniana a tal punto che essa chiese di essere seppellita in mare. Il suo desiderio non venne esaudito e fino al 1947 il suo corpo rimase esposto. Finalmente venne cremato e le ceneri furono sparse nell'oceano.

Forse questa mancanza di rispetto per i morti avrà fine quando i ricercatori cominceranno a studiare i propri antenati o quelli che sarebbero felici di poter considerare come tali. «Ah! Lincoln, un fenotipo interessante», disse la Cann l'anno scorso, quando prese informazioni circa una proposta di ricostruire la sequenza del DNA del grande statista. Gli storici desiderano sapere se la faccia scarna, l'aspetto alto e dinoccolato e la lassità delle articolazioni fossero sintomi della sindrome di Marfan, un'alterazione del tessuto connettivo che può provocare aneurismi potenzialmente fatali. Se il sangue sulle lenzuola del letto di morte di Lincoln mostrasse di contenere il gene responsabile di tale sindrome, i pazienti affetti dalla sindrome di Marfan potrebbero sentirsi orgogliosi di questo illustre precedente e gli storici potrebbero chiedersi se Booth per caso non avesse assassinato un uomo già destinato a morire a breve scadenza.

Ma, per onorare i morti, dobbiamo sapere chi sono stati e da dove sono venuti. Queste domande, che ogni fanciullo si pone e alle quali ogni cultura deve dare risposta, hanno spronato lo sviluppo della scienza moderna. Gli studiosi rinascimentali hanno applicato per primi la ragione e la sperimentazione per ricostruire i testi originali a partire da copie corrotte. Poi i filologi si sono spinti più in là, ricostruendo lingue estinte dalle testimonianze offerte dalle lingue dei discendenti. Questi metodi genealogici, nel contempo scientifici e storici, hanno raggiunto il loro pieno rigoglio nella biologia evoluzionistica, il cui rampollo più giovane, la genetica molecolare, viene oggi in soccorso degli storici.